

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:38:25 ; Search time 30.38 Seconds  
(Without alignments)  
1104.157 Million cell updates/sec

Title: US-09-652-743A-12  
Perfect score: 1559  
Sequence: 1 MKVKKLIPSLVFGALVSP.....NGGPKATAAFVEINTAKAQ 302

ring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5  
747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	1559	100.0	302	22	AAE00016	Bacillus pumilus J
2	1506	96.6	303	22	AAE00014	Bacillus pumilus B
3	1118	71.7	215	22	AAE00196	Bacillus pumilus R
4	1117	71.6	215	22	AAE00197	Bacillus pumilus R
5	1117	71.6	215	22	AAE00200	Bacillus pumilus R
6	1116	71.6	215	22	AAE00199	Bacillus pumilus R
7	1110	71.2	215	22	AAE00198	Bacillus pumilus R
8	435.5	27.9	318	22	AAE00015	Bacillus lichenifo
9	420.5	27.0	316	13	AA23730	Protease Blase. S
10	419.5	26.9	316	22	AAE00011	Bacillus lichenifo
11	417	26.7	314	22	AAE00013	Bacillus lichenifo

12	412.5	26.5	222	22	AAE00075	Bacillus lichenifo
13	411.5	26.4	222	22	AAE00036	Bacillus lichenifo
14	411.5	26.4	222	22	AAE00116	Bacillus lichenifo
15	411.5	26.4	222	22	AAE00118	Bacillus lichenifo
16	410.5	26.3	222	22	AAE00071	Bacillus lichenifo
17	409.5	26.3	222	22	AAE00046	Bacillus lichenifo
18	409.5	26.3	222	22	AAE00048	Bacillus lichenifo
19	409	26.2	221	22	AAE00187	Bacillus lichenifo
20	409	26.2	222	22	AAE00053	Bacillus lichenifo
21	409	26.2	222	22	AAE00190	Bacillus lichenifo
22	408.5	26.2	222	22	AAE00092	Bacillus lichenifo
23	408.5	26.2	222	22	AAE00137	Bacillus lichenifo
24	407.5	26.1	222	22	AAE00047	Bacillus lichenifo
25	407.5	26.1	222	22	AAE00049	Bacillus lichenifo
26	407.5	26.1	222	22	AAE00140	Bacillus lichenifo
27	407	26.1	222	22	AAE00193	Bacillus lichenifo
28	406.5	26.1	222	22	AAE00033	Bacillus lichenifo
29	406.5	26.1	222	22	AAE00076	Bacillus lichenifo
30	406.5	26.1	222	22	AAE00107	Bacillus lichenifo
31	406.5	26.1	222	22	AAE00182	Bacillus lichenifo
32	406.5	26.1	222	22	AAE00205	Bacillus lichenifo
33	406	26.0	222	22	AAE00112	Bacillus lichenifo
34	405.5	26.0	222	12	AA241159	Serine protease SP
35	405.5	26.0	222	13	AA26467	Sequence of a seri
36	405.5	26.0	222	22	AAE00050	Bacillus lichenifo
37	405.5	26.0	222	22	AAE00052	Bacillus lichenifo
38	405.5	26.0	222	22	AAE00073	Bacillus lichenifo
39	405.5	26.0	222	22	AAE00077	Bacillus lichenifo
40	405.5	26.0	222	22	AAE00078	Bacillus lichenifo
41	405.5	26.0	222	22	AAE00108	Bacillus lichenifo
42	405.5	26.0	222	22	AAE00109	Bacillus lichenifo
43	405.5	26.0	222	22	AAE00170	Bacillus lichenifo
44	405.5	26.0	222	22	AAE00172	Bacillus lichenifo
45	405.5	26.0	222	22	AAE00184	Bacillus lichenifo

ALIGNMENTS

RESULT 1

AAE00016  
ID AAE00016 standard; Protein; 302 AA.

XX AC AAE00016;

XX DT 31-MAY-2001 (first entry)

XX DE Bacillus pumilus JA96 RP-II protease.

XX XX Residual protease II; RP-II; additive; cleaning composition; detergent.

XX OS Bacillus pumilus JA96.

XX FH Key Location/Qualifiers

XX FT Peptide 1..25 /label= Signal\_peptide

XX FT Peptide 26..87

XX FT Peptide /label= Pro\_peptide

XX FT Protein /note= "Pro sequence of RP-II protease"

XX FT Protein 88..302

XX FT Protein /note= "Bacillus pumilus JA96 mature RP-II protease"

XX PN WO200116285-A2.

XX PD 08-MAR-2001.

XX PF 31-AUG-2000; 2000WO-DK00476.

XX PR 31-AUG-1999; 99DK-0001212

XX PR 21-OCT-1999; 99DK-0001500.

XX PA (NOVO ) NOVOZYMES AS.

PI Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
PI Flensted Lassen S;  
XX WPI; 2001-226680/23.  
DR N-PSDB; AAD03004.  
XX

XX Novel RP-II type protease and its variants useful as constituents in  
PT detergent compositions, additives and cleaning compositions -  
PS  
PS Claim 1; Page 119-120; 132pp; English.

XX The present sequence is Bacillus pumilus JA96 RP-II (Residual  
XX protease II). RP-II protease is useful as a constituent in additives,  
CC detergent compositions and other cleaning compositions, optionally in  
CC combination with other enzymes such as proteases, lipases, cellulases,  
CC amylases, peroxidases or oxidases. The variants of RP-II have improved  
CC properties such as substrate specificities, catalytic rate, stability,  
CC especially towards the action of proteolytic enzymes and improved  
CC resistance towards peroxidase.

Sequence 302 AA;

Query Match 100.0%; Score 1559; DB 22; Length 302;  
Best Local Similarity 100.0%; Pred. No. 1.5e-116; Indels 0; Gaps 0;  
Matches 302; Conservative 0; Mismatches 0;  
QY 1 MKKKVKKLIPSLVFGALSVPSFAHAASDVLTSYDMVTSDGKVISSADFNHDMKTPSSF 60  
Db 1 mkkvkklipslvfgalsvpsfahaasdsvltsydmvtsgdkvissadfnhdmktpsf 60  
QY 61 DKVDLSTIGEKKVPLTTLTKDFQTVIGDGRKVTNTRVAPYNSIAYITFGSSCT 120  
Db 61 dkvdlsstigeekvpkplttlykdfqtkvvgdgrtkvtntrvapynsiayitfgssct 120  
QY 121 GTLIAPNKLITNGHCYVNTATRSYSKAGSVYPCGNDSTAVNGSANMTEFVPSGYINTGA 180  
Db 121 gtlipnklitngchcyvntatrsyskagsvypcgndstavnsgsanmtefypsgyintga 180  
QY 181 SQYDFAVIKTDNIGNTVGRYSIRQVTLNLTGTTIKISGYPGDKMRSTGKYSQWEMSGPVT 240  
Db 181 sqydfaviktdnigntvgrysirqvtlnltgttikisgypgdkmrstgkysqwemsgpvt 240  
QY 241 REDTNLAYTIDTFSGNSGAMLDQNOQIVGVHAGYSNGTINGGPKATAAFVEFINYAK 300  
Db 241 redtnlaytldtfsgnsgamldqnoqivgvhnagysngtinggpkataafvefinyak 300  
QY 301 AQ 302  
Db 301 aq 302

JT 2  
0014  
AAE00014 standard; Protein; 303 AA.  
XX AAE00014;  
XX AC AAE00014;  
XX DT 31-MAY-2001 (first entry)  
XX DE Bacillus pumilus B032 RP-II protease.  
XX KW Residual protease II; RP-II; additive; cleaning composition; detergent.  
XX OS Bacillus pumilus B032.  
XX PH Key Location/Qualifiers  
FT Peptide 1..26  
FT Peptide /label= Signal\_peptide  
FT Peptide 27..88  
FT Peptide /label= Pro\_peptide  
FT Peptide /note= "Pro sequence of RP-II protease"  
FT Protein 89..303

FT  
XX  
PN  
XX  
PD  
XX  
XX  
PF  
XX  
PR  
XX  
XX  
PA  
XX  
PI  
XX  
XX  
DR  
XX  
PT  
XX  
XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ

WO200116285-A2.  
08-MAR-2001.  
31-AUG-2000; 2000WO-DK00476.  
31-AUG-1999; 99DK-0001212  
20-OCT-1999; 99DK-0001500  
(NOVO ) NOVOZYMES AS.  
Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
Flensted Lassen S;  
WPI; 2001-226680/23.  
N-PSDB; AAD03002.  
Novel RP-II type protease and its variants useful as constituents in  
detergent compositions, additives and cleaning compositions -  
Claim 1; Page 112-113; 132pp; English.

The present sequence is Bacillus pumilus B032 RP-II (Residual  
protease II). RP-II protease is useful as a constituent in additives,  
detergent compositions and other cleaning compositions, optionally in  
combination with other enzymes such as proteases, lipases, cellulases,  
amylases, peroxidases or oxidases. The variants of RP-II have improved  
properties such as substrate specificities, catalytic rate, stability,  
especially towards the action of proteolytic enzymes and improved  
resistance towards peroxidase.

Sequence 303 AA;

Query Match 96.6%; Score 1506; DB 22; Length 303;  
Best Local Similarity 95.7%; Pred. No. 2.6e-112; Indels 0; Gaps 0;  
Matches 289; Conservative 7; Mismatches 6;  
QY 1 MKKKVKKLIPSLVFGALSVPSFAHAASDVLTSYDMVTSDGKVISSADFNHDMKTPSSF 60  
Db 2 mkkvkklipslvfgalsvpsfahaasdsvltsydmvtsgdkvissadfnhdmktpsf 61  
QY 61 DKVDLSTIGEKKVPLTTLTKDFQTVIGDGRKVTNTRVAPYNSIAYITFGSSCT 120  
Db 62 dkvdlsstigeekvpkplskylkdfqtkvvgdgrtkvtntrvapynsiayitfgssct 121  
QY 121 GTLIAPNKLITNGHCYVNTATRSYSKAGSVYPCGNDSTAVNGSANMTEFVPSGYINTGA 180  
Db 122 gtlipnklitngchcyvntatrsyskagsvypcgndstavnsgsanmtefypsgyintga 181  
QY 181 SQYDFAVIKTDNIGNTVGRYSIRQVTLNLTGTTIKISGYPGDKMRSTGKYSQWEMSGPVT 240  
Db 182 sqydfaviktdnigntvgrysirqvtlnltgttikisgypgdkmrstgkysqwemsgpvt 241  
QY 241 REDTNLAYTIDTFSGNSGAMLDQNOQIVGVHAGYSNGTINGGPKATAAFVEFINYAK 300  
Db 242 redtnlaytldtfsgnsgamldqnoqivgvhnagysngtinggpkataafvefinyak 301  
QY 301 AQ 302  
Db 302 aq 303  
RESULT 3  
AAE00196  
ID AAE00196 standard; Protein; 215 AA.  
XX AC AAE00196;  
XX DT 31-MAY-2001 (first entry)  
XX

DE Bacillus pumilus RP-II protease variant (E155A).  
XX Residual protease II; RP-II; additive; cleaning composition; detergent;  
KW mutant; mutein; variant.  
XX Bacillus pumilus JA96.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 155  
FT /note= "Wild type Glu substituted with Ala"  
XX  
XX  
XX WO200116285-A2.  
XX 08-MAR-2001.  
XX  
XX 31-AUG-2000; 2000WO-DK00476.  
XX 31-AUG-1999; 99DK-00012124.  
XX 20-OCT-1999; 99DK-0001500.  
XX  
XX (NOVO ) NOVOZYMES AS.  
XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
XX Flønsted Lassen S;  
XX  
XX WPI; 2001-226680/23.  
XX  
XX Novel RP-II type protease and its variants useful as constituents in  
XX detergent compositions, additives and cleaning compositions -  
XX  
XX Example 3; Page -: 132pp; English.  
XX  
XX The present sequence is Bacillus pumilus (JA96) RP-II protease  
XX variant (E155A).  
XX The patent discloses RP-II (Residual protease II) protein, their  
XX corresponding nucleic acid sequences and variants. RP-II protease  
XX is useful as a constituent in additives, detergent compositions and  
XX other cleaning compositions, optionally in combination with other  
XX enzymes such as proteases, lipases, cellulases, amylases, peroxidases  
XX or oxidases. The variants of RP-II have improved properties such as  
XX substrate specificities, catalytic rate, stability, especially towards  
XX the action of proteolytic enzymes and improved resistance towards  
XX peroxidase.  
XX Note: This sequence is not shown in the specification but is derived  
XX from the mature B. licheniformis RP-II protein [SEQ ID NO: 2] shown in  
XX page 101-102 of the specification (AAE00011). The specification also  
XX refers to following variants: (a) VIF, (b) D7G-T125S+E152G+N182I,  
XX (c) S119A-S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+C185A,  
XX (g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a)  
XX and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in  
XX (d), Gln residue in (e) and (h) do not match with the B. licheniformis  
XX native wild type RP-II protease sequence.  
XX  
XX Sequence 215 AA;  
XX  
XX Query Match 71.7%; Score 1118; DB 2; Length 215;  
XX Best Local Similarity 99.5%; Pred. No. 1.5e-87;  
XX Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX 88 VVIGDGRKVTNTRVAPVNSIAYITFGSSCTGLIAPNKLITNGCHVYNATRSYSAK 147  
XX  
XX 1 vvigddgrkvtntvrvapvnsiayitfgssctgtliapnklitngchvynatrsysak 60  
XX  
XX 148 GSVPFCMNDSTAVNGSANMTEFVPSGINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 207  
XX  
XX 61 gsvpfcmndstavnsgsanmtefvpvsgyintgasqdydfaviktdnigntvgyrsirvgt 120  
XX  
XX 208 NLTGTTIKISGPGKMRSTGKVSQWMSGPVTRDNTLAYTIDTFSGNSGSAMLDQNO 267  
XX  
XX 121 nltgttikisgpgdkmrstgkvsqwmvgpvtradtlayttdtfsgnsgsamlldqno 180

OY 268 QIVGVNAGYSNGTINGGPKATAAFVFEFINYAKAQ 302  
DB 181 qivgvnagysngtinggpkataafvfefinyakaq 215  
RESULT 4  
AAE00197 ID AAE00197 standard; Protein; 215 AA.  
XX AC AAE00197;  
XX 31-MAY-2001 (first entry)  
XX  
XX Bacillus pumilus RP-II protease variant (E155G).  
DE Residual protease II; RP-II; additive; cleaning composition; detergent;  
KW mutant; mutein; variant.  
XX  
XX Bacillus pumilus JA96.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 155 /note= "Wild type Glu substituted with Gly"  
XX  
XX WO200116285-A2.  
XX 08-MAR-2001.  
XX  
XX 31-AUG-2000; 2000WO-DK00476.  
XX  
XX 31-AUG-1999; 99DK-00012124.  
XX 20-OCT-1999; 99DK-0001500.  
XX  
XX (NOVO ) NOVOZYMES AS.  
XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
XX Flønsted Lassen S;  
XX  
XX WPI; 2001-226680/23.  
XX  
XX Novel RP-II type protease and its variants useful as constituents in  
XX detergent compositions, additives and cleaning compositions -  
XX  
XX Example 3; Page -: 132pp; English.  
XX  
XX The present sequence is Bacillus pumilus (JA96) RP-II protease  
XX variant (E155G).  
XX The patent discloses RP-II (Residual protease II) protein, their  
XX corresponding nucleic acid sequences and variants. RP-II protease  
XX is useful as a constituent in additives, detergent compositions and  
XX other cleaning compositions, optionally in combination with other  
XX enzymes such as proteases, lipases, cellulases, amylases, peroxidases  
XX or oxidases. The variants of RP-II have improved properties such as  
XX substrate specificities, catalytic rate, stability, especially towards  
XX the action of proteolytic enzymes and improved resistance towards  
XX peroxidase.  
XX Note: This sequence is not shown in the specification but is derived  
XX from the mature B. licheniformis RP-II protein [SEQ ID NO: 2] shown in  
XX page 101-102 of the specification (AAE00011). The specification also  
XX refers to following variants: (a) VIF, (b) D7G-T125S+E152G+N182I,  
XX (c) S119A-S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+C185A,  
XX (g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a)  
XX and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in  
XX (d), Gln residue in (e) and (h) do not match with the B. licheniformis  
XX native wild type RP-II protease sequence.  
XX  
XX Sequence 215 AA;  
XX  
XX Query Match 71.6%; Score 1117; DB 22; Length 215;  
XX Best Local Similarity 99.5%; Pred. No. 1.8e-81;  
XX Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 88 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGTLLIAPNKKILTNHCHVYNTATRSYSAK 147  
 |||||  
 Db 1 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGTLLIAPNKKILTNHCHVYNTATRSYSAK 60  
 |||||  
 QY 148 GSVYPCMNDSSTAVNGSANMTEFVYVPSGYINTGASQYDFAVIKTDNIGNTGVYRSIRQVT 207  
 |||||  
 Db 61 GSVYPCMNDSSTAVNGSANMTEFVYVPSGYINTGASQYDFAVIKTDNIGNTGVYRSIRQVT 120  
 |||||  
 QY 208 NLGTCTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYYYTIDTFSGNSGSAMLDDNQ 267  
 |||||  
 Db 121 NLGTCTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYYYTIDTFSGNSGSAMLDDNQ 180  
 |||||  
 QY 268 QIVGVHNAGYSNGTNGGPKATAAFVEFINYAKAQ 302  
 |||||  
 Db 181 QIVGVHNAGYSNGTNGGPKATAAFVEFINYAKAQ 215  
 |||||

## RESULT 5

AAE00200  
 AAE00200 standard; Protein: 215 AA.

AAE00200;

DT 31-MAY-2001 (first entry)

Bacillus pumilus RP-II (first entry)

Residual protease II: RP-II; additive; cleaning composition; detergent;  
 mutant; mutein; variant.

Bacillus pumilus JA96.

Synthetic.

Key Location/Qualifiers

FT Misc-difference 156

FT /note- "Wild type Asp substituted with Gly"

WO200116285-A2

08-MAR-2001

31-AUG-2000; 2000WO-DK00476.

31-AUG-1999; 99DK-0001212 ✓  
 20-OCT-1999; 99DK-0001500 ✓

(NOVO ) NOVOZYMES AS.

Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
 Flensted Lassen S;

WPI; 2001-226680/23.

Novel RP-II type protease and its variants useful as constituents in  
 detergent compositions, additives and cleaning compositions -

Example 3; Page -: 132pp; English.

The present sequence is Bacillus pumilus (JA96) RP-II protease

variant (D156G).

The patent discloses RP-II (Residual protease II) protein, their  
 corresponding nucleic acid sequences and variants. RP-II protease  
 is useful as a constituent in additives, detergent compositions and  
 other cleaning compositions, optionally in combination with other  
 enzymes such as proteases, lipases, cellulases, amylases, peroxidases  
 or oxidases. The variants of RP-II have improved properties such as  
 substrate specificities, catalytic rate, stability, especially towards  
 the action of proteolytic enzymes and improved resistance towards  
 peroxidase.

Note: This sequence is not shown in the specification but is derived  
 from the mature B. licheniformis RPII protein [SEQ ID NO: 2] shown in  
 page 101-102 of the specification (AAE00011). The specification also

CC refers to following variants: (a) VIF, (b) D7G+T125S+E152G+N182I,  
 CC (c) S119A+S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+Cl85A,  
 CC (g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a)  
 CC and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in  
 CC (d), Gln residue in (e) and (h) do not match with the B. licheniformis  
 CC native wild type RPII protease sequence.

SQ Sequence 215 AA;

Query Match 71.6%; Score 1117; DB 22; Length 215;  
 Best Local Similarity 99.5%; Pred. No. 1.8e-81;  
 Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 88 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGTLLIAPNKKILTNHCHVYNTATRSYSAK 147  
 |||||  
 Db 1 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGTLLIAPNKKILTNHCHVYNTATRSYSAK 60  
 |||||

QY 148 GSVYPCMNDSSTAVNGSANMTEFVYVPSGYINTGASQYDFAVIKTDNIGNTGVYRSIRQVT 207  
 |||||  
 Db 61 GSVYPCMNDSSTAVNGSANMTEFVYVPSGYINTGASQYDFAVIKTDNIGNTGVYRSIRQVT 120  
 |||||

QY 208 NLGTCTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYYYTIDTFSGNSGSAMLDDNQ 267  
 |||||  
 Db 121 NLGTCTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYYYTIDTFSGNSGSAMLDDNQ 180  
 |||||

QY 268 QIVGVHNAGYSNGTNGGPKATAAFVEFINYAKAQ 302  
 |||||

Db 181 QIVGVHNAGYSNGTNGGPKATAAFVEFINYAKAQ 215  
 |||||

## RESULT 6

AAE00199  
 ID AAE00199 standard; Protein: 215 AA.

AC AAE00199;

DT 31-MAY-2001 (first entry)

Bacillus pumilus RP-II protease variant (D156A).

Residual protease II: RP-II; additive; cleaning composition; detergent;  
 mutant; mutein; variant.

Bacillus pumilus JA96.

Synthetic.

Key Location/Qualifiers

FT Misc-difference 156

FT /note- "Wild type Asp substituted with Ala"

WO200116285-A2

08-MAR-2001

31-AUG-2000; 2000WO-DK00476.

31-AUG-1999; 99DK-0001212 ✓  
 20-OCT-1999; 99DK-0001500 ✓

(NOVO ) NOVOZYMES AS.

Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
 Flensted Lassen S;

WPI; 2001-226680/23.

Novel RP-II type protease and its variants useful as constituents in  
 detergent compositions, additives and cleaning compositions -

Example 3; Page -: 132pp; English.

The present sequence is Bacillus pumilus (JA96) RP-II protease

CC variant (D156A).  
 CC The patent discloses RP-II (Residual protease II) protein, their  
 CC corresponding nucleic acid sequences and variants. RP-II protease  
 CC is useful as a constituent in additives, detergent compositions and  
 CC other cleaning compositions, optionally in combination with other  
 CC enzymes such as proteases, lipases, cellulases, amylases, peroxidases  
 CC or oxidases. The variants of RP-II have improved properties such as  
 CC substrate specificities, catalytic rate, stability, especially towards  
 CC the action of proteolytic enzymes and improved resistance towards  
 CC peroxidase.  
 CC Note: This sequence is not shown in the specification but is derived  
 CC from the mature B. licheniformis RPII protein [SEQ ID NO: 2] shown in  
 CC page 101-102 of the specification (AAE00011). The specification also  
 CC refers to following variants: (a) V1F, (b) D7G+T125S+E152G+N182I,  
 CC (c) S119A+S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+C185A,  
 CC (g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a)  
 CC and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in  
 CC (d), Gln residue in (e) and (h) do not match with the B. licheniformis  
 CC native wild type RPII protease sequence.  
 XX  
 SQ Sequence 215 AA;

Query Match 71.6%; Score 1116; DB 22; Length 215;  
 Best Local Similarity 99.5%; Pred. No. 2.2e-81;  
 Matches 214; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 88 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAK 147  
 DB 1 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAK 60  
 QY 148 GSVYPCMDNDSTAVNGSANMTEFVPSGYINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 207  
 DB 61 GSVYPCMDNDSTAVNGSANMTEFVPSGYINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 120  
 QY 208 NLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYTYTIDTFSGNSGSMILDDNQ 267  
 DB 121 NLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYTYTIDTFSGNSGSMILDDNQ 180  
 QY 268 QIVGVHNAGYSNGTINGGPKATAAFVEFINYAKAQ 302  
 DB 181 QIVGVHNAGYSNGTINGGPKATAAFVEFINYAKAQ 215

RESULT 7  
 AAE00198  
 AAE00198 standard; Protein: 215 AA.  
 AAE00198;  
 31-MAY-2001 (first entry)  
 Bacillus pumilus RP-II protease variant (E155A+D156A).  
 Residual protease II: RP-II; additive; cleaning composition; detergent;  
 mutant; mutein; variant.  
 Bacillus pumilus JA96.  
 Synthetic.

Key Location/Qualifiers  
 FT Misc-difference 155  
 FT /note= "Wild type Glu substituted with Ala"  
 FT Misc-difference 156  
 FT /note= "Wild type Asp substituted with Ala"

WO200116285A2  
 08-MAR-2001.  
 31-AUG-2000; 2000WO-DK00476.  
 31-AUG-1999; 99DK-0001212.

PR 20-OCT-1999; 99DK-0001500.  
 XX (NOVO ) NOVOZYMES AS.  
 XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
 PI Flensted Lassen S;  
 XX WPI: 2001-226680/23.  
 PT Novel RP-II type protease and its variants useful as constituents in  
 PT detergent compositions, additives and cleaning compositions  
 PS Example 3; Page -: 132pp; English.  
 XX The present sequence is Bacillus pumilus (JA96) RP-II protease  
 CC variant (E155A+D156A).  
 CC The patent discloses RP-II (Residual protease II) protein, their  
 CC corresponding nucleic acid sequences and variants. RP-II protease  
 CC is useful as a constituent in additives, detergent compositions and  
 CC other cleaning compositions, optionally in combination with other  
 CC enzymes such as proteases, lipases, cellulases, amylases, peroxidases  
 CC or oxidases. The variants of RP-II have improved properties such as  
 CC substrate specificities, catalytic rate, stability, especially towards  
 CC the action of proteolytic enzymes and improved resistance towards  
 CC peroxidase.  
 CC Note: This sequence is not shown in the specification but is derived  
 CC from the mature B. licheniformis RPII protein [SEQ ID NO: 2] shown in  
 CC page 101-102 of the specification (AAE00011). The specification also  
 CC refers to following variants: (a) V1F, (b) D7G+T125S+E152G+N182I,  
 CC (c) S119A+S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+C185A,  
 CC (g) E152A+V144L and (h) E104K+Q204R. However the positions of Val in (a)  
 CC and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in  
 CC (d), Gln residue in (e) and (h) do not match with the B. licheniformis  
 CC native wild type RPII protease sequence.  
 XX  
 SQ Sequence 215 AA;

Query Match 71.2%; Score 1110; DB 22; Length 215;  
 Best Local Similarity 99.1%; Pred. No. 6.7e-81;  
 Matches 213; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 88 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAK 147  
 DB 1 VVIGDGRKVTNTRVAPYNSIAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAK 60  
 QY 148 GSVYPCMDNDSTAVNGSANMTEFVPSGYINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 207  
 DB 61 GSVYPCMDNDSTAVNGSANMTEFVPSGYINTGASQYDFAVIKTDNIGNTVGYRSIRQVT 120  
 QY 208 NLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYTYTIDTFSGNSGSMILDDNQ 267  
 DB 121 NLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYTYTIDTFSGNSGSMILDDNQ 180  
 QY 268 QIVGVHNAGYSNGTINGGPKATAAFVEFINYAKAQ 302  
 DB 181 QIVGVHNAGYSNGTINGGPKATAAFVEFINYAKAQ 215

RESULT 8  
 AAE00015  
 ID AAE00015 standard; Protein: 318 AA.  
 XX  
 AC AAE00015;  
 XX  
 DT 31-MAY-2001 (first entry)  
 XX  
 DE Bacillus licheniformis CDJ31 RP-II protease.  
 XX Residual protease II: RP-II; additive; cleaning composition; detergent.  
 OS Bacillus licheniformis CDJ31.



D	b		121	sslgsgctgwmigpkvtatagchidydtssgsfagtattvpgprngtstypysgvskstryflps	180
Q	y		174	GYYNTCASQYDPAVIKTDNTNIGTGVGRSIRQVT-NLTGTTIKISGYPGDKMRSTGKVSVQ	232
D	b		181	gw-rsgntnydygailesepigntcvgfygysyttsslvgttvtisgpgdktagt----	q 235
Q	y		233	WEMSGPVTRFDNLAYITIDFSCNSGSAMLQNQQ-----IVCVINAG-YNGNTIN	283
D	b		236	wdhsqglaisetykiqyamdtlyggsgspvfedqssrtnscgpcslavhtnvgvygssyn	295
Q	y		284	GGPKAT	289
D	b		296	rgcrlt	301
<hr/>					
		'LT	10		
		0011			
A	C		AAE00011	standard; Protein; 316 AA.	
X	X		AAE00011;		
X	X		31-MAY-2001	(first entry)	
D	E		Bacillus licheniformis (BLC)	RP-II protease.	
X	X		Residual protease II; RP-II;	additive; cleaning composition; detergent.	
K	W		Bacillus licheniformis.		
X	X		Key	Location/Qualifiers	
F	H		Peptide	1..31	
F	T		/label=	Signal_peptide	
P	T		Peptide	32..94	
F	T		/label=	Pro_peptide	
F	T		/note=	"Pro sequence of RP-II protease"	
F	T		Protein	95..316	
F	T		/note=	"Bacillus licheniformis mature RP-II protease"	
X	N		WO200116285-A2.		
X	P		08-MAR-2001.		
X	P		31-AUG-2000;	2000WO-DK00476 ✓	
X	P		31-AUG-1999;	99DK-0001212 ✓	
P	R		20-OCT-1999;	99DK-0001500 ✓	
<hr/>					
		(NOVO )	NOVOZYMES AS.		
P	I		Norreagaard-madsen M,	Rahbek Ostergaard P,	Voge Christensen CB;
X	R		Flensted Lassen S;		
D	R		WPI:	2001-226680/23.	
X	P		N-PSDB;	AAD02999.	
X	P		Novel RP-II type protease and its variants useful as constituents in	detergent compositions, additives and cleaning compositions -	
X	P		Claim 1;	Page 101-102; 132pp;	English.
X	C		The present sequence is Bacillus licheniformis (BLC) RP-II (Residual	protease II). RP-II protease is useful as a constituent in additionally,	
X	C		detergent compositions and other cleaning compositions, optionally in	combination with other enzymes such as proteases, lipases, cellulases,	
X	C		amylases, peroxidases or oxidases. The variants of RP-II have improved	properties such as substrate specificities, catalytic rate, stability,	
X	C		especially towards the action of proteolytic enzymes and improved	resistance towards peroxidase.	
X	S		Sequence	316 AA;	

Query Match 26.9%; Score 419.5; DB 22; Length 316;

Best Local Similarity 33.9%; Pred. No. 1.5e-25;  
Matches 106; Conservative 49; Mismatches 117; Indels 41; Gaps 11;

QY 2 KVKK-LIPSLVFGALSV---PSFAHAASDSVLTS-----YD-MVTS DGKVIS 46  
||| : || : | : | : | : | : | : | : | : | : | : | :  
Db 5 ksvkrglitglisylgmhpaaqapshcpvssdpsykaetsvydphiksdqgly 64  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
QY 47 SAPFNDMKTPSSFDKVDLLSSTIGEKVKPLTTYLKDPOTKVVIIGDDGRTKVTNTRVAPY 106  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
Db 65 skaftgvgknetkek-----aeekspaka---pysiksvisgsddrtrvtnttaypy 113  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
QY 107 NSTAYITFGSSCTGTLIAPNKILTNCHCYNTATRSYSAKGSVYPGMNDSTAVNGSANM 166  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
Db 114 raivhassigsctgmwigpkvtataghciydlssgsfagtatvspgrngtspygsvks 173  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
QY 167 TEFVVPSCGYINTCASOYDFAVIKTDTNIGHTNVGVRSIRQVT-NLTGTTIKISGYPGDKMR 225  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
Db 174 tryfipsgw-rsntnydygaiealsepigtvgyfsyttsslvgttvisgypgdkta 232  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
QY 226 STKVSQWMSGPVTREDTNLAYYYITDFTSGNSGSAMLQNQQ-----IVGVHNAG- 276  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
Db 233 gt----qwqhsghplaisetyklyamtyggsgspvfegssrtncsgpcslavhtngv 288  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
QY 277 YSNCTINGCPKAT 289  
||| : ||| : | : | : | : | : | : | : | : | : | : | :  
Db 289 yggssynrgtrit 301  
||| : ||| : | : | : | : | : | : | : | : | : | : | :

RESULT 11  
AAE00013  
ID AAE00013 standard; Protein: 314 AA.  
XX AAE00013;  
AC AC  
XX XX  
DT 31-MAY-2001 (first entry)  
XX XX  
DE Bacillus licheniformis AC116 RP-II protease.  
XX XX  
KW Residual protease II; RP-II: additive; cleaning composition; detergent.  
XX XX  
OS Bacillus licheniformis.  
XX XX  
FH Key Location/Qualifiers  
FT Peptide 1..29  
FT FT /label= Signal\_peptide  
FT Peptide 30..92  
FT FT /label= Pro-peptide  
FT FT /note= "Pro sequence of RP-II protease"  
FT FT 93..314  
FT FT /note= "Bacillus licheniformis AC116 mature RP-II  
FT FT protease".  
XX PN W0200116285-A2.  
XX XX  
08-MAR-2001  
XX XX  
31-AUG-2000; 2000WO-DK00476.  
XX XX  
31-AUG-1999; 99DK-0001212 ✓  
PR PR 20-OCT-1999; 99DK-0001500. ✓  
XX XX  
(NOVO ) NOVOZYMES AS.  
XX XX  
PI Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
PI PI Flønsted Lassen S;  
XX XX  
WIPI; 2001-226680/23.  
DR DR N-PSDB; AAD03001.  
XX XX  
PT Novel RP-II type protease and its variants useful as constituents in  
PT PT detergent compositions, additives and cleaning compositions -  
XX XX Claim 1; Page 109-110; 132pp; English.

Claim 1; Page 109-110; 132pp; English.







Db 121 slvgtvtisgpgdktagt----qwhsgplaisetykqyamdyggsgspvfegss 176  
QY 268 Q-----IVGVHAG-YSNGTINGGPKAT 289  
Db 177 srtnsgpcslavhtngvyggssynrgtrlt 207

RESULT 15 -  
AAE00118  
ID AAE00118 standard; Protein; 222 AA.  
XX AC AAE00118;  
XX DT 31-MAY-2001 (first entry)  
XX DE Bacillus licheniformis RP-II protease variant (T128K).  
XX FW Residual protease II; RP-II; additive; cleaning composition; detergent;  
mutant; mutein; variant.  
Bacillus licheniformis.  
Synthetic.

XX Key Location/Qualifiers  
XX FT Misc-difference 128  
XX FT /note= "Wild type Thr substituted with Lys"  
XX PN W0200116285-A2.  
XX PD 08-MAR-2001.  
XX PF 31-AUG-2000; 2000WO-DK00476.  
XX PR 31-AUG-1999; 99DK-0001212.6 ✓  
XX PR 20-OCT-1999; 99DK-0001500.  
XX PA (NOVO ) NOVOZYMES AS.  
XX PI Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;  
XX PI Flensted Lassen S;  
XX DR WPI; 2001-226680/23.  
XX PT Novel RP-II type protease and its variants useful as constituents in  
XX PT detergent compositions, additives and cleaning compositions  
XX PS Example 3; Page -; 132pp; English.  
XX CC The present sequence is Bacillus licheniformis (BLC) RP-II protease  
variant (T128K).  
XX CC The patent discloses RP-II (Residual protease II) protein, their  
corresponding nucleic acid sequences and variants. RP-II protease  
is useful as a constituent in additives, detergent compositions and  
other cleaning compositions, optionally in combination with other  
enzymes such as proteases, lipases, cellulases, amylases, peroxidases  
or oxidases. The variants of RP-II have improved properties such as  
substrate specificities, catalytic rate, stability, especially towards  
the action of proteolytic enzymes and improved resistance towards  
peroxidase.  
XX CC Note: This sequence is not shown in the specification but is derived  
from the mature B. licheniformis RPII protein [SEQ ID NO: 2] shown in  
page 101-102 of the specification (AAE00011). The specification also  
refers to following variants: (a) V1F, (b) D7G+T125S+E152G+N182I,  
XX CC (c) S119A+S144T+Q157R+S175I, (d) H141A, (e) E152A+Q122R, (f) N182T+C185A,  
XX CC (g) E152A+V144L and (h) E104K+Q204R. However the positions of val in (a)  
XX CC and (g), Asn residue in (b) and (f), Ser residues in (c), His residue in  
XX CC (d), Gln residue in (e) and (h) do not match with the B. licheniformis  
native wild type RPII protease sequence.  
XX SQ Sequence 222 AA;

Query Match 26.4%; Score 411.5; DB 22; Length 222;  
Best Local Similarity 39.8%; Pred. No. 4e-25;  
Matches 84; Conservative 36; Mismatches 76; Indels 15; Gaps 5;  
QY 89 VIGDDGRKTKVTRVAPYNSIAYITFGSSCTGLIAPNKILTNCHVCYNTATRSYSAKG 148  
Db 2 vigsddtrvtunttaypyralvhisssigscgwmigpaktvatagchciydissysfagta 61  
QY 149 SVYFGMNDSTAVNGSANNTEFYVPSGYINTGASQYDFAVIKTD7NIGNTVGYRSIROVT- 207  
Db 62 tvspgrngtspygsvkstryfipsgw-rsgntnydygaielsepignvtvgfygysyts 120  
QY 208 NLGCTTIKISGPGDKWRSTGKVSQWEMSGPVTREDTNLAYYYTIDTFSGNSGSAMLQNOQ 267  
Db 121 slvgttvkisgpgdktagt----qwhsgplaisetykqyamdyggsgspvfegss 176  
QY 268 Q-----IVGVHAG-YSNGTINGGPKAT 289  
Db 177 srtnsgpcslavhtngvyggssynrgtrlt 207

Search completed: June 21, 2002, 18:41:56  
Job time: 211 sec

Copyright (c) 1993 - 2000 Compucon  
GENCORE version 4.3

Result No.	Score	Query Match	Length	DB ID	Description
2	1501	96.3	303	2 Q9EXR9	Q9EXR9 bacillus in
3	341	21.9	358	16 Q931E7	Q931E7 staphylococ
3	254	16.3	364	16 Q98G17	Q98G17 rhizobium l
4	227.5	14.6	639	16 Q92LB9	Q92LB9 rhizobium m
5	214	13.7	316	2 Q9F6G1	Q9F6G1 staphylococ
6	202.5	13.0	188	16 Q99R42	Q99R42 staphylococ
7	192.5	12.3	357	2 Q04186	Q04186 staphylococ
8	189.5	12.2	342	16 Q99V45	Q99V45 staphylococ
9	189	12.1	217	2 Q9AJX0	Q9AJX0 staphylococ
10	187.5	12.0	271	16 Q987W6	Q987W6 rhizobium l
11	179.5	11.5	284	2 Q47809	Q47809 enterococcu
12	174.5	11.2	240	2 Q9KH50	Q9KH50 staphylococ
13	173.5	11.1	240	16 Q953781	Q953781 staphylococ
14	161	10.3	238	2 Q9FD07	Q9FD07 staphylococ
15	156	10.0	239	2 Q9KH49	Q9KH49 staphylococ
16	156	10.0	330	16 Q9K1E3	Q9K1E3 vibrio chol

1500  
Judy Ph. City  
Kit No. 1

2 MKKVKMLLP SLLVFGALSPPSFAHATSDPSVLTSDYDMVTSDGKV I SSSDFHN DTKSPSSF 61

Native

QY 61 DKVDDLSSTIGEKVKPLTTLTKDFQFQWVIGDGRTRKVTNTRVAPYNSIAYITFGSSCT 120  
Db 62 DKVDDLSSTIGEKVKPLTTLTKDFQFQWVIGDGRTRKVTNTRVAPYNSIAYITFGSSCT 121  
QY 121 GTLIAPNKILTNHCHVCYNTATRSYSAKSGVYPCGMDSTAVNGSANNTEFYVPSGYINTCA 180  
Db 122 GTLIAPNKILTNHCHVCYNTATRSYSAKSGVYPCGMDSTAVNGSANNTEFYVPSGYINTCA 181  
QY 181 SQYDFAVIKTDNIGNTVGRSIRQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVT 240  
Db 182 SQYDFAVIKTDNIGNTVGRSIRQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGSVT 241  
QY 241 REDTNLAYTIDTFSGNSGSMLDQNOQIVGVHNGYSGTNGGPKATAAFVEFYNYAK 300  
Db 242 REDTNLAYTIDTFSGNSGSMLDQNOQIVGVHNGYSGTNGGPKATAAFVEFYNYAK 301  
QY 301 AQ 302  
Db 302 AQ 303  
T 2  
AC Q931E7 PRELIMINARY; PRT; 358 AA.  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DE HYPOTHETICAL PROTEIN SAV2596.  
GN SAV2596.  
OS Staphylococcus aureus (strain Mu50).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=158878;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus.";  
RL Lancet 357:1225-1240(2001).  
DR EMBL: AP003365; BAB58758.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 358 AA; 41086 MW; 5AB5B04202583B08 CRC64;

ry Match 21.9%; Score 341; DB 16; Length 358;  
t Local Similarity 32.6%; Pred. No. 7.2e-15;  
atches 91; Conservative 51; Mismatches 111; Indels 26; Gaps 9;  
QY 30 VLTSY---DMVTSKGVISSAFHNDKTPS-SFDKVDLSSTIGEKVKPLTTLTKDFQ 85  
Db 15 VIVSSYGESEFVLADNNPIVNSGSDNNFSGTIERNQNELAEVGDYSTPL----- 66  
QY 86 TKVVGDDGRTRKVTNTRVAPYNS--IAYITFGSSC---TGTLIAPNKLITNGHCYNTA 140  
Db 67 TERVFGKQDQRTVNNLTQRPYKGTGTVLLNMTFSNNRYKGTGTGTMICKDIIVLTAHNVYSD 126  
QY 141 TRSYSAKSGVYPCGMDSTAVNGSANNTEFYVPSGYINTCASOYDFAVIKTDNIGNTVGY 200  
Db 127 DKGWAKIDVYAGVNGQTITIGKAFSHKFFVSKTINNAPTREDIAIKLSNGLNKGTY 186  
QY 201 RSRQVNTLT-GTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAYITIDTFSGNSG 259  
Db 187 LTLN--THLSKGENIISGFGPKDSN----RQYKGKLGESFDENEMYTYVDTFSGSG 240  
QY 260 SAMLDQNOQIVGVHNGYSGTNGGPKATAAFVEFYNY 298

Db 241 SAIRDSKNNIIGHVAYGRYNNH--NSGVRINDKLVDYNY 277  
RESULT 3  
Q98G17 PRELIMINARY; PRT; 364 AA.  
AC Q98G17  
DT 01-OCT-2001 (TRENBLrel. 18, Created)  
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
DE GLUTAMYL ENDOPEPTIDASE (EC 3.4.21.19).  
GN MLL3306.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
DR EMBL: AP003001; BAB50229.1; --  
DR InterPro; IPR00126; Ser\_proteas\_v8.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00839; V8PROTEASE.  
DR SMART; SM00020; Tryp\_SPC; 1.  
KW Hydrolase; Complete proteome.  
SQ SEQUENCE 364 AA; 39034 MW; 13BE653270E7CDD3 CRC64;

Query Match 16.3%; Score 254; DB 16; Length 364;  
Best Local Similarity 30.3%; Pred. No. 3.7e-09;  
Matches 70; Conservative 38; Mismatches 95; Indels 28; Gaps 7;  
QY 89 VIGDGRTRKVTNTRVAPYNSIAYI-----TFGSSCTGTLIAPNKLITNGHCYNTATR 142  
Db 136 VFGPDDREQVRNTKTYPFSAIGYLEAKSAKTSYGSCSATLIGPRTVLTAAHCLYSHEDK 195  
QY 143 SYSAGSVYPCGMDSTAVN---GSANMTEFYVPSGYINTGASQY-----DFAVIKTDN 193  
Db 196 DMLSEYLFVPLNGSTADDAPEGAFYTESAYVLOGFIDNYQGYGVSIVPMDLGIITLAKD 255  
QY 194 IGMTVGRSIRQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTREDTNLAY--YTI 251  
Db 256 VGTNGLWGLYANDLGDFTANLVGYPGDKPMGT-----MVKASCEVHAENIAPEYFQYDC 311  
QY 252 DTFSGNSGSMI-----DQNOQIVGVHNGYSGTNGGPKATAAFVEFYNY 297  
Db 312 DTFPGSSGSVVAYDTKSKQRIITGVNVA--ESPDAVTLNLAANVQWIN 360  
RESULT 4  
Q92LB9 PRELIMINARY; PRT; 639 AA.  
AC Q92LB9  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DE PUTATIVE PROTEASE TRANSMEMBRANE PROTEIN.  
GN SMC03769.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;

```

RN      SEQUENCE FROM N.A.
RP      STRAIN-1021;
RX      MEDLINE=21368234; PubMed=11474104;
RX      Gallibert F., Finan T.M., Long S.R., Puchler A., Abola P., Ampe F.,
RA      Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA      Boutry M., Bower L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA      Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA      Gloux S., Godrie T., Goffeau A., Golding B., Gouz J., Gurjal M.,
RA      Hernandez-Lucas I., Hong A., Hulzar L., Hyman R.W., Jones T., Kahn D.,
RA      Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA      Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA      Ramsperger U., Surzycki R., Thebault P., Vandenbol M.,
RA      Vorhoefer F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RA      *The composite genome of the legume symbiont Sinorhizobium meliloti.*;
RA      Science 293:668-672(2001).
RA      EMBL: AL591793; CAC47733.1; -.
RA      Complete proteome.
RA      SEQUENCE 639 AA; 69813 MW; 320E86D3050DFB67 CRC64;

Query Match          14.6%; Score 227.5; DB 16; Length 639;
Best Local Similarity 30.3%; Pred. No. 3 9e-07;
Matches 63; Conservative 31; Mismatches 83; Indels 31; Gaps 8;

Qy      89 VIGDGRKTKYNTVPYNSIAYITFGSS-----CTGTLIAPNKILNGHCYVN-TATR 142
Db      412 VIGEDTRAAVLDTGTFPSRAIVQILFETRAREQHLCTGTLVSPNVLTAACHIHSGTRTG 471

Qy      143 SYSKAGSYPCGNDSTAVNGSANNTEFVPSGYNTGASQ-----YDFAVIKTDNIGNVTV 198
Db      472 EPFONFRIRIIEPGNIGAAPPFCGLGVAGSVLAGTASATTDQSRYYDGLGAIKNCNIGDPT 531

Qy      199 GYRSIROVTNLGTITIKISGYPGDKMRSTGK--VSQ-----WEMSGPVTREDTNLAYVT 250
Db      532 GWLGVRTIGNDEAIDTVVQGYAADR-APTGRQWSEDKLRILWOLKG-----FYQ 580

Qy      251 IDTFSGNSGSAML--DQNOQIVGVHNAG 276
Db      581 NDTFGGTSGAPVFAKDSDTDLIGVHTNG 608

RESULT 5
Q9FBG1 PRELIMINARY; PRT; 316 AA.
AC
Q9FBG1;
01-MAR-2001 (TrEMBLrel. 16, Created)
01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GLUTAMYL ENDOPEPTIDASE.
GN
OS Staphylococcus warneri.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1292;
RN
RN SEQUENCE FROM N.A.
RA Kakikawa M.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Kodaira K.I.;
RA *Characterization of the gene encoding glutamyl endopeptidase of
RT Staphylococcus warneri M.*;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ293885; CAC06168.1; -.
DR MEROPS: S01.269; -.
DR InterPro: IPR000126; Ser.proteas_v8.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.

```

[illegible]

Db 2 AIIKNSNLGNTGTLN--THISGENIEISGPGKSDN----ROYKGGKLESTDE 55

QY 245 NLAITYTIDTFSNGSAMLQNOQIVGVHAGYSNGTNGGPKAPAAFEVFINY 298  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 56 NEMYTYVTDFSGSGSADRNSKNIIGVHAYGRYNH--NSGVRINDLKLDYINY 107

RESULT 7

Q04186 PRELIMINARY; PRT; 357 AA.

AC Q04186;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE GLUTAMIC ACID SPECIFIC PROTEASE PREPROPEPTIDE (EC 3.4.21.19).

OS Staphylococcus aureus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Staphylococcus.

OX NCBI\_TaxID=1280;

[1]

SEQUENCE FROM N.A.  
Yoshikawa K., Tsuzuki H., Fujiwara T., Nakamura E., Iwamoto H.,  
Matsumoto K., Shin M., Yoshida N., Teraoka H.;  
\*Purification, characterization and gene cloning of a novel glutamic  
acid-specific endopeptidase from staphylococcus aureus atcc 12600.\*;  
RL Biochim. Biophys. Acta 1121:221-228(1991).

DR EMBL; D00730; BAA00630.1; .

DR MEROPS; S01.269; .

DR InterPro; IPR000126; Ser\_proteas\_v8.

DR Pfam; PF00089; trypsin.

DR PRINTS; PR00839; trypsin; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00672; V8\_HIS; 1.

DR PROSITE; PS00673; V8\_SER; 1.

KW Hydrolase; Protease; Serine protease.

FT CHAIN 69 357 POTENTIAL.

SQ SEQUENCE 357 AA; 38651 MW; 58AA9A4E371E2577 CRC64;

Query Match 12.3%; Score 192.5; DB 2; Length 357;  
Best Local Similarity 23.8%; Pred. No. 3.8e-05;  
Matches 77; Conservative 47; Mismatches 127; Indels 73; Gaps 14;

QY 3 KVKKLIPSLVFGALSVPSFAHAASDVLTSDYDMVTSKGVISSADFNHMKTPSSFDK 62  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 7 KVSLFVATLTATLVSSPAANALSSKAMDNHPQQTQTD-----KQTP---K 51

QY 63 VDDLSTIGEKKVPLTLYLKDFQTKVIGDGRKVTNTRVAPYNSIAYITF----GGSS 118  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 52 IQK-----GGNLKPLE--QREANVILPNDRHQITDTTNGHYAPVTYIOVEAPTGTPI 103  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

119 CTGTLIAPNKILTNHGVVNTATRSYSAKGSVYP-GMNDSTAVNGSANNTEFYVPSGYIN 177  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

104 ASGVVVGKDTLTNKHVVVDATHGDPHALK--AFPSAINODNYPNGGFTAEOITKYSG--- 158

QY 178 TGASQYDFAVIKTDTNIGN-----TVGYRSIQVTNLTGTIKISGPGDKMRST 227  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 159 ----EGDLAIVFSPNEQNKHIGEVVVKPATMSNNAETQVQ---NITVGYPGDKPVAT 210  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 228 KGVSWHEMSGPVTREDTNLAYTIDTFSNGSAMLQNOQIVGVHAGYSNGTNGGPK 287  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 211 ----MWESKGIYTLKGEAMQYDLSTTGGNSGSPVFNKNEKNEVIGIHMGVGN-QFNG---- 262  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 288 ATAFAF-----VEFINYA 299  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 263 --AVFINENVRNFKONIEDINFA 284

RESULT 8

Q09V45 PRELIMINARY; PRT; 342 AA.

ID Q09V45

AC Q09V45;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE SERINE PROTEASE, V8 PROTEASE, GLUTAMYL ENDOPEPTIDASE.

GN SSPA OR SA0901 OR SAV1048.

OS Staphylococcus aureus (strain N315), and .

OS Staphylococcus aureus (strain Mu50).

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Staphylococcus.

OX NCBI\_TaxID=158879; 158878;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);

RA MEDLINE=21311952; PubMed=11418146;

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
\*Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus.\*;  
RL Lancet 357:1225-1240(2001).

CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.

DR EMBL; AP003132; BAB42146.1; .

DR EMBL; AP003361; BAB57210.1; .

DR InterPro; IPR000126; Ser\_proteas\_v8.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00839; V8PROTEASE.

DR SMART; SM00020; Tryp\_SPC; 1.

DR PROSITE; PS00240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00672; V8\_HIS; 1.

DR PROSITE; PS00673; V8\_SER; 1.

KW Complete proteome; Hydrolase; Protease; Serine protease.

FT CHAIN 69 357 POTENTIAL.

SQ SEQUENCE 342 AA; 36977 MW; 5AEF42DCE01C4B24 CRC64;

Query Match 12.2%; Score 189.5; DB 16; Length 342;  
Best Local Similarity 24.2%; Pred. No. 5.7e-05;  
Matches 74; Conservative 45; Mismatches 126; Indels 61; Gaps 12;

QY 3 KVKKLIPSLVFGALSVPSFAHAASDVLTSDYDMVTSKGVISSADFNHMKTPSSFDK 62  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 7 KVSLFVATLTATLVSSPAANALS-----SKAMDNHPQQTQSSKQO 48  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 63 VDDLSTIGEKKVPLTLYLKDFQTKVIGDGRKVTNTRVAPYNSIAYITF----GGSS 118  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 49 TPKIKK--GGNLKPLE--QREHANVILPNDRHQITDTTNGHYAPVTYIOVEAPTGTPI 103  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 119 CTGTLIAPNKILTNHGVVNTATRSYSAKGSVYP-GMNDSTAVNGSANNTEFYVPSGYIN 177  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 104 ASGVVVGKDTLTNKHVVVDATHGDPHALK--AFPSAINODNYPNGGFTAEOITKYSG--- 158

QY 178 TGASQYDFAVIKTDTNIGN-----TVGYRSIQVTNLTGTIKISGPGDKMRST 227  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 159 ----EGDLAIVFSPNEQNKHIGEVVVKPATMSNNAETQVQ---NITVGYPGDKPVAT 210  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 228 KGVSWHEMSGPVTREDTNLAYTIDTFSNGSAMLQNOQIVGVHAGYSNGTNGGPK 287  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 211 ----MWESKGIYTLKGEAMQYDLSTTGGNSGSPVFNKNEKNEVIGIHMGVGN-EPNG---- 262  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 288 ATAFAF 293  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 263 --AVFI 266

RESULT 9

Q09AJX0 PRELIMINARY; PRT; 217 AA.

ID Q09AJX0

AC Q09AJX0;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

```
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE EXTRACELLULAR SERINE PROTEINASE PRECURSOR (FRAGMENT).
GN ESP.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6746;
RA Dubin G.;
RT "Staphylococcus epidermidis extracellular serine proteinase.";
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ305145; CAC27157.1; -.
DR HSSP; P09331; 1EXF.
DR InterPro; IPR001254; Ser_proteas_V8.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT NON_TER 1 1
FT SIGNAL <1 1 POTENTIAL.
FT CHAIN 2 217 EXTRACELLULAR SERINE PROTEINASE.
SQ SEQUENCE 217 AA; 23667 MW; FB9B886D4538BB7 CRC64;

Query Match 12.18; Score 189; DB 2; Length 217;
Best Local Similarity 28.4%; Pred. No. 3.6e-05;
Matches 59; Conservative 32; Mismatches 83; Indels 34; Gaps 8;

Qy 88 VVIGDGRKTVNTRVAPYNSIAYITF---GGSSCTCTLIAPNKILTNHGVNTATRS 143
Dl : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 VILPNNRHQIFNTTOGHYDAVSFIYIPIDGYMSGGVVGVEILLTNKHVV----- 54

Qy 144 YSAGK-----SVYP-GMNDSTAVNGSANMTEFYVPSGYNTGASQYDFAVIKTDN----- 193
Dl : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 NGAKGNPRNISVHPSAKNENDYPNGKFVQEIIYPG-----NSDLAILRVSPNEHNQ 107

Qy 194 -IGNTVGYSIRQVTNL-TGTTIKISGYPGDKMRSTGKVSQWMSGPPVTREDTNLAYITI 251
Dl : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 HIGQVVKPRTSSNTDTRINENITVTGPGDKPLAT-----MWESVGKVVYIGBELRYDL 163

Qy 252 DTFSGNSGSAMLDNQIQIVGVHNGYSN 279
Dl : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 STVGNSGSPVFNKQVIGIHYGGVDN 191

RESULT 10
Q987W6 PRELIMINARY; PRT; 271 AA.
AC Q987W6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE GLUTAMIC ACID SPECIFIC ENDOPEPTIDASE.
GN MLL6878.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
```

```
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003010; BAB53084.1; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000126; Ser_proteas_V8.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
KW Complete proteome.
SQ SEQUENCE 271 AA; 30385 MW; 37659307C63D0D9C CRC64;

Query Match 12.08; Score 187.5; DB 16; Length 271;
Best Local Similarity 27.6%; Pred. No. 5.8e-05;
Matches 66; Conservative 42; Mismatches 98; Indels 33; Gaps 11;

Qy 81 LKDFQTKVIG-DGRTKVTNTRVAPYNSIAYI--TFPG---SSCTGTLIAPNKILTNHGH 134
Dl : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 LADEADYAVVGPTDGRARVTHTRFPHSAVCHIERDFGDRMTGCTAFLLISPTLLTAAH 73

Qy 135 CVYNTATR-----SYSAGSVYPCMNDSA-VNGSANMTEFYVPSGYNTGASQYDFAVI 188
Dl : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 CITSPIROLGLPLNLAIRVTPGRASRDARPFQWQAKQHWNPYRRRPSGLHDVGLI 133

Qy 189 KTDNIGNTVGYSI-----FOVTNLGT-TIKISGYPGDKMRSTGKVSQWMSGPPVTR 242
Dl : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 134 ELERPFSPGCHFQWSPNRODLERLNRLLHISGYFADKPDGT---QWEHSELDRI 189

Qy 243 DTNLAYITIDTFSGNSGSAMLDNQOQ-----IVGVHNG---YSNGTIN---GGPKATA 290
Dl : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 TERQLFYSVDTCPGHSAGPVVTHROQAGPPVVIATAGPRPHSGGANGCRPGVPLA 248

RESULT 11
Q47809 PRELIMINARY; PRT; 284 AA.
AC Q47809;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE STAPHYLOCOCCAL SERINE PROTEINASE HOMOLOGUE.
GN SPRE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OGI-10;
RA Su Y.A., Clewell D.B.;
RT "A gene (spre) downstream of gele of Enterococcus faecalis OGI-10
RT resembles serine proteinase determinant of Staphylococcus aureus
RT strain V8.";
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z12296; CAA78168.1; -.
DR MEROPS; S01.00B; -.
DR InterPro; IPR000126; Ser_proteas_V8.
DR PRINTS; PR00839; V8PROTEASE.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; UNKNOWN_1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 284 AA; 31063 MW; D4F0312BEE778415 CRC64;

Query Match 11.5%; Score 179.5; DB 2; Length 284;
Best Local Similarity 25.5%; Pred. No. 0.00021;
Matches 68; Conservative 40; Mismatches 96; Indels 63; Gaps 11;
```

QY 66 LSTTGEKVKPLTTLTKDQTKVVGDDGRTKVTNTRVAPYNSI-----AYITFG 115  
DB 27 LSAEAEYIVPAESHRSQ-KRSLDDPRRQEVADTTEAPFASIGRIITSPASKPGYISLG 85  
QY 116 GSSCTGTLIAPNKILTNCHCV---YNTATRSYSAGSV--YPCMDNSTAVNGSANMTEP- 169  
DB 86 ----TGFVVGTTVTNTNNHVAESFNKAKVLNPNKADDAWFYGRGDSATPGKFKRVIDVA 141  
QY 170 YVPSGYINTGASQYDFAVIKTDNTIGNTVGYRSIR-----QVTNLTGTTI 214  
DB 142 FSPNA-----DIAVV-----TVGKQNDPRDGPDELGEILLTPFVLKKKFESSDTHV 184  
QY 215 KISGPGDKMRSTGKVSQHE-----MSGPVTREDTNNLAYITDTSGNSGSAAMLQNOOI 269  
DB 185 TISGYPGEKNH-----TOWSHENDLFTSNFTDLENPLLFYDIDTTGGQSGSPIYNAQFEV 239  
QY 270 VGVHNGYSGNTINGPKATAAFVEFI 296  
DB 240 VGVHSGGKIQGNHGORLNEVNYNFI 266

T 12  
30

ID Q9KH50 PRELIMINARY; PRT; 240 AA.  
AC Q9KH50;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SERINE PROTEASE SPLB.  
GN SPLB.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=8325-4;  
RX MEDLINE=21116966; PubMed=11179322;  
RA Reed S.B., Wesson C.A., Liou L.E., Trumble W.R., Schlievert P.M.,  
RA Bohach G.A., Bayles K.W.;  
RT "Molecular Characterization of a Novel Staphylococcus aureus Serine  
RT Protease Operon";  
RL Infect. Immun. 69:1521-1527(2001).  
DR EMBL; AF271715; AAF97926.1; -;  
DR MEROPS; S01.282; -;  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00839; V8PROTEASE.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00672; V8\_HIS; 1.  
KW Complete proteome; Hydrolase; Serine protease.  
SQ SEQUENCE 240 AA; 26096 MW; 00A3349F1FE5D52E CRC64;

Query Match 11.2%; Score 174.5; DB 2; Length 240;  
Best Local Similarity 28.4%; Pred. No. 0.00036;  
Matches 61; Conservative 31; Mismatches 78; Indels 45; Gaps 11;  
QY 78 TTYLKDFQKVVIGDGRKVTNTRVAPYNSIAYIFGSSCTGTLIAPNKILTNCHCVY 137  
DB 24 TTLVEEVQ-OTAKAENNVTKIQDTNIFPYTGV--VAF--KSATGFFVVGKNTILTNKHVSK 78  
QY 138 N-----TATRS-----YSAGSV-YPGMNDSTAVNGSANMTEFFVPSGY-INTCA 180  
DB 79 NYKVGDRITAHPNDSKNGGIYSIKKIINYPGKEDYSVQVEERAIE-RGPGGFNFNDV 137  
QY 181 SOYDFAVTKDNTNIGTVGYRSIRQVTNLTGTITIKISYPGDKMRSTGKVSQWEMSGPVT 240  
DB 138 TPEKYAA-----GAKAGER-----IKVIGYPHP---YKNKYVLYESTGPGVM 175  
QY 241 REDTNLAYITDTTFSGNSGSAAMLQNOQIVGVHNA 275

DB 176 SVEGSSIVYSAHTESGSGSPVLNSNNELVGIHFA 210  
RESULT 13  
Q53781 PRELIMINARY; PRT; 240 AA.  
ID Q53781;  
AC Q53781;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NOVEL ANTIGEN (SERINE PROTEASE SPLB).  
GN SPLB OR SA1630 OR SAV1812.  
OS Staphylococcus aureus, (strain N315), and  
OS Staphylococcus aureus (strain N315), and  
OS Staphylococcus aureus (strain Mu50).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=1280, 158879, 158878;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.aureus; STRAIN=FDA #485 (ABD ENT+);  
RA Rieneck K., Renneberg J., Gutschik E., Diamant M., Svenson M.,  
RA Bendtzen K.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus";  
RL Lancet 357:1225-1240(2001).  
CC -! SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
DR EMBL; U60589; AAB03613.1; -;  
DR EMBL; AP003135; BAB42898.1; -;  
DR EMBL; AP003363; BAB57974.1; -;  
DR MEROPS; S01.282;  
DR InterPro; IPR00126; Ser\_proteas\_v8.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00839; V8PROTEASE.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00672; V8\_HIS; 1.  
KW Complete proteome; Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 240 AA; 26140 MW; 45EF0B5A51E06DA7 CRC64;

Query Match 11.1%; Score 173.5; DB 16; Length 240;  
Best Local Similarity 27.9%; Pred. No. 0.00042;  
Matches 60; Conservative 32; Mismatches 78; Indels 45; Gaps 11;  
QY 78 TTYLKDFQKVVIGDGRKVTNTRVAPYNSIAYIFGSSCTGTLIAPNKILTNCHCVY 137  
DB 24 TTLVEEVQ-OTAKAENNVTKIQDTNIFPYTGV--VAF--KSATGFFVVGKNTILTNKHVSK 78  
QY 138 N-----TATRS-----YSAGSV-YPGMNDSTAVNGSANMTEFFVPSGY-INTCA 180  
DB 79 NYKVGDRITAHPNDSKNGGIYSIKKIINYPGKEDYSVQVEERAIE-RGPGGFNFNDV 137  
QY 181 SOYDFAVTKDNTNIGTVGYRSIRQVTNLTGTITIKISYPGDKMRSTGKVSQWEMSGPVT 240  
DB 138 TPEKYAA-----GAKAGER-----IKVIGYPHP---YKNKYVLYESTGPGVM 175  
QY 241 REDTNLAYITDTTFSGNSGSAAMLQNOQIVGVHNA 275  
DB 176 SVEGSSIVYSAHTESGSGSPVLNSNNELVGIHFA 210



```

RT      *Molecular Characterization of a Novel Staphylococcus aureus Serine
RT      Protease Operon.;
RL      Infect. Immun. 69:1521-1527(2001).
DR      EMBL; AF271715; AAF97927.1; -
DR      MEROPS; S01.283; -
DR      InterPro; IPR001940; Protease2C.
DR      InterPro; IPR001126; Ser_proteas_V8.
DR      InterPro; IPR001254; Trypsin.
DR      Pfam; PF00089; trypsin; 1.
DR      PRINTS; PR00834; PROTEASES2C.
DR      PRINTS; PR00839; V8PROTEASE.
DR      PROSITE; PS50240; TRYPSIN_DOM; 1.
DR      PROSITE; PS00672; V8_HIS; 1.
DR      Hydrolase; Protease; Serine protease.
SQ      SEQUENCE 239 AA; 26098 MW; BA797E14786B2DBA CRC64;

      Query Match      10.0%;      Score 156;      DB 2;      Length 239;
      Best Local Similarity 25.9%;      Pred. No. 0.0058;
      Matches 57;      Conservative 29;      Mismatches 88;      Indels 46;      Gaps 11;

QY      96  TKYTNTRVAPYNSIAYITFGSSCTGLIAPNKILTNHCV--YNTATR----- 142
Db      41  TVQKDTNIPYNGV--VSF--KDATGFVICKNTIITNKHVSKDYKVGDRITAHPNGDKGN 96

QY      143 --SYSAKG-SVYPGMNDSTAVNGSANNTEFYVPSGY-INTGASQYDFAVIKTDNIGNTV 198
Db      97  GGIYKIKSISDYPGEDISIVNIEEQAVE-RGPKGFNFENVOAFNFA---KDAKVDD-- 150

QY      199  GYRSIRQVTNLTTGIKISYPGDKMRSTCKVSWQMSGPGVTREDDTNLAYVITDIFSGNS 258
Db      151 -----KIKVIGYP---LPAQNSFKQFESTGTGIKRIKDKNILNFDAYIEFGNS 193

QY      259  GSAMLDNQOQIVGVHNNAGYS--NGTINGGPKATAAFAVEFI 296
Db      194  GSPVLNSNNEVIGVYGGICKIGSEYNGAVFTTPIKDFI 233

Search completed: June 21, 2002, 18:45:41
Job time: 220 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:40:36 ; Search time 17.42 Seconds  
(without alignments)  
1665.843 Million cell updates/sec

Title: US-09-652-743A-12  
Perfect score: 1559  
Sequence: 1 MKKVKKLIPLSLVFGALSPV.....NGGPKATAAFVEFIFYAKAQ 302

```

ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```

```

Searched:      283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :
PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	420.5	27.0	316	2	A45134	endopeptidase (EC
2	280.5	18.0	313	2	A35122	metalloproteinase
3	202.5	13.0	188	2	E90065	hypothetical prote
4	192.5	12.3	357	2	S21758	glutamic acid-spec
5	189.5	12.2	336	1	PRSASK	glutamyl endopepti
6	189.5	12.2	342	2	G98873	hypothetical prote
7	182.5	11.7	278	2	A40282	probable peptidas
8	179.5	11.5	284	2	S25140	serine proteinase
9	173.5	11.1	240	2	C99967	serine proteinase
10	170.5	10.9	218	2	E97915	choline binding pr
11	156	10.0	330	2	B82415	probable serine pr
12	155.5	10.0	280	1	PRSAEA	epidermolytic toxi
13	154	9.9	273	2	E85765	hypothetical prote
14	153.5	9.8	273	2	H84915	putative protease
15	153.5	9.8	285	2	C95045	choline binding pr
16	150	9.6	239	2	B99967	serine proteinase
17	141.5	9.1	239	2	H89966	serine proteinase
18	139.5	8.9	235	2	D99967	serine proteinase
19	137.5	8.8	239	2	A89967	serine proteinase
20	135	8.7	273	2	A80682	probable secreted
21	133.5	8.6	168	2	D90066	hypothetical prote
22	133	8.5	277	1	PRSAEB	epidermolytic toxi
23	130	8.3	1275	2	T33369	hypothetical prote
24	128.5	8.2	272	2	JC4170	trypsin-like prote
25	125	8.0	1651	2	JC1340	outer membrane pro
26	124	8.0	598	2	A92222	hypothetical prote
27	121	7.8	397	2	H70789	probable serine pr
28	120.5	7.7	488	2	B69825	cell wall-binding
29	120	7.7	254	1	TRWV3Y	trypsin-like prote

## ALIGNMENTS

RESULT 1  
A45134  
C:Species: Bacillus licheniformis  
C:Accession: A45134; S23078  
C:Date: 10-Jun-1993  
C:Revision: 18-Nov-1994 #text\_change 15-Oct-1999  
C:Accession: A45134; S23078  
C:Author: Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, T.  
J Biol Chem 267: 23782-23788, 1992  
A:Title: Purification, characterization, cloning, and expression of a glutamic acid  
A:Reference number: A45134; MUID:93054737  
A:Accession: A45134  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-316 <KAK>  
A:Cross-references: GB:d10060; NID:g216263; PIDN:BAA00949.1; PID:d1001415; PID:g216263  
A:Experimental source: ATCC 14580  
A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBIIP:118785)  
R:Sandberg, K.  
R:Svendsen, I.; Breddam, K.  
Eur J Biochem 204, 165-171, 1992  
A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase  
A:Reference number: S23078; MUID:92155199

Query Match	27.0%	Score	420.5	DB	2	Length	316		
Best Local Similarity	33.0%	Pred. No.	2.1e-22						
Matches	101	Conservative	55	Mismatches	121	Indels	29		
						Gaps	10		
QY	5	KKLIPSLIVFGALSWPSPA	---	HAASDSVLTSYDMDWTS	DG	-	KVSISSADFHNDMKT	---	56
		:	:	:	:	:	:	:	
DbB	4	KKSVRGLITGLIGISYGLMHPA	-	QAAPSPHTPVSSDP	SPSYKAETSVTYDNIK	SDQY	61		
QY	57	---	PSSEFKVDL	STICEGVKPLUT	TVLKDFQTKVWIGD	GRKTVNTRVAP	NSIAYIT	113	
		:	:	:	:	:	:	:	
DbB	62	GLYSKFTGCTGVNET	-	KEKAEEKSPAKAPYSIK	YSGSDRTRVTN	TYATPR	IVHIS	120	
QY	114	FGSSCTCTLIAPKNLTHGCVNT	ATRSYSAKGSYVPGMND	STAVNGSANNTE	FYVPS	173			
		:	:	:	:	:	:	:	
DbB	121	SSIGSCTGMWIGPKTVATAG	HCIIYDTSSGSFAGT	ATVSPGRNGTSYVPGSVKSTR	FIPS	180			
QY	174	GYINTGASQYDPAVIKTD	TNIGTVGYSRINQVT	-	NLTGTTIKISGYPGDKMR	STOKVSQ	232		
		:	:	:	:	:	:	:	
DbB	181	GW-RSGNTNYDYGAT	ELSEPIGNTVG	YFGYSYTTSSLVGT	TVTITSGYPGDKTAGT	---	Q	235	
QY	233	WEMSGPVTREDTNLAYIT	DTDFSSNGSSAML	DQNQQ	-----	IVGVHNAG	-YSNCTIN	283	
		:       :	:       :	:       :	:       :	:       :	:       :	:       :	
DbB	236	WQHSQGP	IAISETYKIQYAMDTYGGQSG	VPFEQSSSRNCS	NGSCGPCSLAVHTNGYVGGSSY	295			

[illegible]

QY 63 VDDLSTIGEKKVPLTYLTKDFQTKVIGDDGRTKVTNTRVAPYNSIAYITF----GGSS 118  
Db 52 IQK-----GGNLKPLE---QREANVILPNDRHQITDTTNGHYAPVTYIQVEAPTGTFI 103  
QY 119 CTGTLIAPNKILTNCHVCYNTATRSYSAKGSYVP-GMNDSTAVNGSANMTEFYVPSGYIN 177  
Db 104 ASGVVVGKDTLLTNKHVVVDATHGDPHALK--AFPSAINODNYPNGGFTAEQITKYSYSG--- 158  
QY 178 TGAQYDFAVIKTDNIGN-----TVGYRSIROVTNLTGTITIKISGYPGDKMRST 227  
Db 159 -----EGDLAIVKFSNEQNKHIGEVVVKPATMSNNAETQVNO-----NITVTGYPGDKPVAT 210  
QY 228 GKVSQWEMSGPVTREDTINLAYITIDTFSGNSGSAMLQDQOQIVGVHNGYSGNCTINGGPK 287  
Db 211 -----MWESKGKITYLKGEAMQYDLSTTGGNSGSPVFNEKNEVIGIHWGVPN-QFNG--- 262  
288 ATAAP-----VEFINYA 299  
263 --AVFINVRNPLKONIEDINFA 284

## RESULT 5

## PRASK

glutamyl endopeptidase (EC 3.4.21.19) precursor - Staphylococcus aureus

N:Alternate names: staphylococcal serine proteinase

C:Species: Staphylococcus aureus

C:Date: 04-Dec-1986 #sequence\_revision 30-Jun-1991 #text\_change 18-Jun-1999

C:Accession: A26812; A00966

R:Carmona, C.; Gray, G.L.

Nucleic Acids Res. 15, 6757, 1987

A:Title: Nucleotide sequence of the serine protease gene of Staphylococcus aureus, strain

A:Reference number: A26812; MUID:87316953

A:Accession: A26812

A:Molecule type: DNA

A:Residues: 1-336 <C>

A:Cross-references: EMBL:Y00356; NID:946686; PIDN:CAA68434.1; PID:946687

A:Experimental source: strain V8

R:Drapeau, G.R.

Can. J. Biochem. 56, 534-544, 1978

A:Title: The primary structure of staphylococcal protease.

A:Reference number: A23824; MUID:78212487

A:Accession: A00966

A:Molecule type: protein

A:Residues: 69-108;110-124;126-144,'D',146-192,'T',194-228,'N',230-258,'Q',260,'D',262-2

A:Experimental source: strain V8

C:Comment: This extracellular proteolytic enzyme preferentially cleaves the peptide bond

argument: This enzyme may be distantly related to the trypsin-type serine proteinases,

perfamily: staphylococcal serine proteinase

eywords: hydrolase; serine proteinase

F:69-336/Product: staphylococcal serine proteinase #status experimental <ACT>

F:119,161/Active site: His, Asp #status predicted

F:237/Active site: Ser #status experimental

Query Match 12.2%; Score 189.5; DB 1; Length 336;  
Best Local Similarity 24.2%; Pred. No. 4.3e-06;  
Matches 74; Conservative 45; Mismatches 126; Indels 61; Gaps 12;

QY 3 KVKKLIPSLVFGALSVPSFAHAASDVLTSDYDMVTSKGVISSADFNDKMTPTSSFDK 62  
Db 7 KVSLLFVATLTATLVSSPAANALS-----SKAMDNHPQQTSSKQO 48

QY 63 VDDLSTIGEKKVPLTYLTKDFQTKVIGDDGRTKVTNTRVAPYNSIAYITF----GGSS 118  
Db 49 TPKIQR--GGNLKPLE---QREANVILPNDRHQITDTTNGHYAPVTYIQVEAPTGTFI 103

QY 119 CTGTLIAPNKILTNCHVCYNTATRSYSAKGSYVP-GMNDSTAVNGSANMTEFYVPSGYIN 177  
Db 104 ASGVVVGKDTLLTNKHVVVDATHGDPHALK--AFPSAINODNYPNGGFTAEQITKYSYSG--- 158

QY 178 TGAQYDFAVIKTDNIGN-----TVGYRSIROVTNLTGTITIKISGYPGDKMRST 227  
Db 159 -----EGDLAIVKFSNEQNKHIGEVVVKPATMSNNAETQVNO-----NITVTGYPGDKPVAT 210

Db 159 -----EGDLAIVKFSNEQNKHIGEVVVKPATMSNNAETQVNO-----NITVTGYPGDKPVAT 210  
QY 228 GKVSQWEMSGPVTREDTINLAYITIDTFSGNSGSAMLQDQOQIVGVHNGYSGNCTINGGPK 287  
Db 211 -----MWESKGKITYLKGEAMQYDLSTTGGNSGSPVFNEKNEVIGIHWGVPN-EFNG--- 262  
QY 288 ATAAPV 293  
Db 263 --AVFI 266

## RESULT 6

## G89873

hypothetical protein sspA [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: G89873

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, J.

ma, A.; Mizutani-Oli, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimi,

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: G89873

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-342 <KUR>

A:Cross-references: GB:BA000018; PID:gl37008050; PIDN:BA842146.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: sspA

C:Superfamily: staphylococcal serine proteinase

Query Match 12.2%; Score 189.5; DB 2; Length 342;  
Best Local Similarity 24.2%; Pred. No. 4.4e-06;  
Matches 74; Conservative 45; Mismatches 126; Indels 61; Gaps 12;

QY 3 KVKKLIPSLVFGALSVPSFAHAASDVLTSDYDMVTSKGVISSADFNDKMTPTSSFDK 62  
Db 7 KVSLLFVATLTATLVSSPAANALS-----SKAMDNHPQQTSSKQO 48

QY 63 VDDLSTIGEKKVPLTYLTKDFQTKVIGDDGRTKVTNTRVAPYNSIAYITF----GGSS 118  
Db 49 TPKIQR--GGNLKPLE---QREANVILPNDRHQITDTTNGHYAPVTYIQVEAPTGTFI 103

QY 119 CTGTLIAPNKILTNCHVCYNTATRSYSAKGSYVP-GMNDSTAVNGSANMTEFYVPSGYIN 177  
Db 104 ASGVVVGKDTLLTNKHVVVDATHGDPHALK--AFPSAINODNYPNGGFTAEQITKYSYSG--- 158

QY 178 TGAQYDFAVIKTDNIGN-----TVGYRSIROVTNLTGTITIKISGYPGDKMRST 227  
Db 159 -----EGDLAIVKFSNEQNKHIGEVVVKPATMSNNAETQVNO-----NITVTGYPGDKPVAT 210

QY 228 GKVSQWEMSGPVTREDTINLAYITIDTFSGNSGSAMLQDQOQIVGVHNGYSGNCTINGGPK 287  
Db 211 -----MWESKGKITYLKGEAMQYDLSTTGGNSGSPVFNEKNEVIGIHWGVPN-EFNG--- 262

QY 288 ATAAPV 293  
Db 263 --AVFI 266

## RESULT 7

## AH0282

probable peptidase YPO2319 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001

C:Accession: AH0282

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentis,

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougl,

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; B;

Nature 413, 523-527, 2001

PAGE 047 : 00

Db	3	KIDNTLQYPYSTAMVLSYKYGVADQMNVGSGSANFIKDNVLIIAAHNYYR---HDYK	59
QY	147	KGS---VYPMGDNSTAVNGSANMTEFYVPSGYINTC---ASQYDFAVIKTDNIGNTVGY	200
Db	60	EADDIVLPAVPSQELFGKIKVKEVYIKERPNLSKDAREYDLALLILEKPIGAKLCT	119
QY	201	RSI-ROVTNLGTFTIKISGYPGDKMRSTCKVSO-WPMMSGPVTRDTNLAYYITDIFSGNS	258
Db	120	LGLPTSQKNLTGITVITITVPSYNF---KIQMYTKDKQVLSDDGMFLDYQVDTLEGSS	175
QY	259	GSAMLDQNOQIVGVHNAGTNSGTINGGPKATAAFVEFI	296
Db	176	GSTVYDASHRVGVHITLGDGANOISAVKLNERNLPFI	213

"ILT 11  
15  
able serine proteinase VCA0803 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
...pecies: Vibrio cholerae  
C : Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
R : Accession: B82415  
R : Heidelberg, J.F.; Eisen, J.A.; Neilson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.  
l., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Query Match 10.0%; Score 156; DB 2; Length 330;  
Best Local Similarity 28.0%; Pred. No. 0.00096;  
Matches 70; Conservative 27; Mismatches 89; Indels 64; Gaps 16;

150	VYPMND-STAVNGS---	ANMTEFYVPSGYINTGASQY--DFAVIK--TDTNIGTVG--	199
	: : : : :	: : : : :	
	: : : : :	: : : : :	
85	VVPOLEDSQFPNGVQLAAAEFYYPDNYVDSAVYWPNDIAIIKLESDLNVSFVGVL	144	
	: : : : :	: : : : :	
	: : : : :	: : : : :	
200	-----YRSIRO-----	VTNLGTGTTIKI-----SGYPGDMRSTGKVS	231
	: : : : :	: : : : :	
145	NSSINNSYDENTGYKAIGHGVNNGVAGSTRLETTLTFTVPATCSAYTGANL-GPGHVC	203	
	: : : : :	: : : : :	
232	QWEMSGPVRTEDTNLAYTIDTFSNGSGAML-DQNOQLVGVHNNAGYSNGTNGGP--KA	288	
	: : : : :	: : : : :	
204	----FTGP-----	QIGSYRNSTCSGDSGPGVYWDSCSGYVQIGTISFGPSTC-GNPALPV	253
	: : : : :	: : : : :	
289	TAAAFEVFINY	298	
	: : : : :	: : : : :	
254	TSVFTEVSDY	263	
	: : : : :	: : : : :	

RESULT 12  
PRSAEA  
epidermolytic toxin A precursor - Staphylococcus aureus  
C:Species: Staphylococcus aureus  
N:Alternate names: exfoliative toxin A  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 18-Jun-1999  
C:Accession: A26680: C26680: A26679: S11541: A46569  
R:Lee, C.Y.; Schmidt, J.J.; Johnson-Winegar, A.D.; Spero, L.; Iandolo, J.J.  
J. Bacteriol. 169, 3904-3909, 1987  
A:Title: Sequence determination and comparison of the exfoliative toxin A and toxin B genes

A:Reference number: A26680; MUID:87307959  
A:Accession: A26680  
A:Molecule type: DNA  
A:Residues: 1-280 <LEE>  
A:Cross-references: GB:M17347; NID:gl53005; PIDN:AAA26625.1; PID:gl53006  
A:Experimental source: strain UT0002  
A:Accession: C26680  
A:Molecule type: protein  
A:Residues: 39-82; 209-248 <LEE2>  
A:Note: the carboxyl-terminal residue of the mature protein is Lys  
R:O'Toole, P.W.; Foster, T.J.  
J. Bacteriol. 169, 3910-3915, 1987  
A:Title: Nucleotide sequence of the epidermolytic toxin A gene of *Staphylococcus*  
A:Reference number: A26679; MUID:87307960  
A:Accession: A26679  
A:Molecule type: DNA  
A:Residues: 1-280 <OTO>  
A:Cross-references: GB:M17357; NID:gl53007; PIDN:AAA26626.1; PID:gl53008  
A:Experimental source: strain TC16  
R:Bailey, C.J.; Smith, T.P.  
Biochem. J. 269, 535-537, 1990  
A:Title: The reactive serine residue of epidermolytic toxin A.  
A:Reference number: S11541; MUID:90351379  
A:Accession: S11541  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 224-240 <BAI>  
R:Sakurai, S.; Suzuki, H.; Kondo, I.  
J. Gen. Microbiol. 134, 711-717, 1988  
A:Title: DNA sequencing of the eta gene coding for staphylococcal exfoliative toxin  
A:Reference number: A46569; MUID:89036121  
A:Accession: A46569  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2, 1', 4-249, 'R', 251-254, 'I', 256-280 <SAK>  
A:Cross-references: GB:M20371  
A:Experimental source: strain ZM  
A:Note: the authors translated the codon ATT for residue 3 as Asn, CAA for residue  
55 as Leu  
C:Genetics:  
A:Gene: eta  
C:Superfamily: staphylococcal serine proteinase  
C:Keywords: scalded skin syndrome; toxin  
F:1-38/Domain: signal sequence #status <SIG>  
F:39-280/Product: epidermolytic toxin A #status predicted <MAT>

Qy 289 TAAFFVEFINYAK

Db 265 ----VGIGNVK 272

RESULT 13

E85765

hypothetical protein 22592 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: E85765

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Lin, A.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E85765

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <STO>

A:Cross-references: GB:AE005174; NID:g12515576; PIDN:AA656585.1; GSPDB:GN00145; UWGP:225

A:Experimental source: strain O157:H7, substrain EDL933

A:Keywords: protease

e: 22592

Query Match 9.8%; Score 154; DB 2; Length 273;

Best Local Similarity 23.6%; Pred. No. 0.001;

Matches 57; Conservative 40; Mismatches 91; Indels 54; Gaps 10;

QY 65 DLSSTIGERKVPKLTTLKDFQTKWIGDGRKVTNTRVAPYNSIAYI-TFGSSCTGTL 123

Db 25 DVAKSANDEVSL-----FFGHDRVPVNDTQSPWDVAGQLETASGNLCTATL 73

QY 124 IAPNKLITNGHCYVN-----TATRSYSAK-----GSVYPMNDSTAVNGS 163

Db 74 IAPNALTAGHCLLTPPKGKADKAVLRVSNKGLWRYDIHDIIEGRVDTLGRKLKADGD 133

QY 164 ANTEFYVPSGYINTCASQYDFA-VIKTDTNIGNTV-----GYRSIROVTNLTGTTIKI 216

Db 134 G-----WIVP-----PAAQPMDFGLVLRNPPSGITPLPFEFGDKAALTAALKAAGRKVTQ 184

QY 217 SGYPGDKMRSTGKVSQWEMSGPVTRDTNLAAYTTIDTFSGNSGSAML-----DONQOIVGVH 273

Db 185 AGYPEDHLOTLYSHQNCVETG---WAQTSVMHQCDTLPGDSSPLMLHTDDGWLIGVQ 241

QY 274 NA 275

Db 242 SS 243

RESULT 14

E85765

active protease ydgD precursor (EC 3.4.21.-) [similarity] - Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 31-Dec-2001

C:Accession: H64915

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: H64915

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-273 <BLAT>

A:Cross-references: GB:AE000255; GB:U00096; NID:g1787875; PIDN:AA674670.1; PID:g1787882;

A:Experimental source: strain K-12, substrain MG1655

C:Keywords: hydrolase; serine proteinase

Query Match 9.8%; Score 153.5; DB 2; Length 273;

Best Local Similarity 24.6%; Pred. No. 0.001;

Matches 55; Conservative 37; Mismatches 89; Indels 43; Gaps 9;

QY 83 DFQTKVWIGDGRKVTNTRVAPYNSIAYI-TFGSSCTGTLIAPNKLITNGHCYVN---- 138

Db 32 DEVSTLFFGHDRVPVNDTQSPWDVAGQLETASGNLCTATLAPNLAITLTPPK 91

QY 139 -----TATRSYSAK-----GSVYPMNDSTAVNGSANMTEFYVPSGYINTGAS 181

Db 92 GRADKAVALLRFVSNKGLWRYEIHDIIEGRVDTLGRKLKADGDG-----WIVP-----PAA 142

QY 182 QYDFA-VIKTDTNIGNTV-----GYRSIROVTNLTGTTIKISGYPGDKMRSTGKVSQWE 234

Db 143 PWFDFGLVLRNPPSGITPLPFEFGDKAALTAALKAAGRKVTQAGYPEDHLDLYSHQNC 202

QY 235 MSSPVTRDTNLAAYTTIDTFSGNSGSAML-----DONQOIVGVHNA 275

Db 203 VTG---WAQTSVMHQCDTLPGDSSPLMLHTDDGWLIGVQSS 243

RESULT 15

C95045

choline binding protein G [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: C95045

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S. on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtz nson, T.; Hickey, E.K.; Holt, I.E. Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Mor A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumonia A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: C95045

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-285 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK74556.1; PID:g14971860; GSPDB:GN00164; T A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0390

Query Match 9.8%; Score 153.5; DB 2; Length 285;

Best Local Similarity 27.0%; Pred. No. 0.0012;

Matches 50; Conservative 26; Mismatches 94; Indels 15; Gaps 6;

QY 120 TGTLIAPNKLITNGHCYVNTRATRSYSAGS---VYPMNDSTAVNGSANMTEFYVPSGYI 176

Db 20 SANFIKDNVLIATAAHNYR---HDYKKEADDIYVLPVAVSPSOEPFGKIKVKEVYLKEFR 76

QY 177 NTG---ASOYDFAVIKTDNIGNTVGYSI-RQVTNLTGTTIKISGYPGDKMRSTGKVSQ 232

Db 77 NLNSKDAREYDLALLILEEPGAKLGLTSLPTSQKSLTGTITVTITGYSYNF----KIHO 132

QY 233 -WEMSGPVTRDTNLAAYTTIDTFSGNSGSAMLDOQOIVGVHNAVYSNCTINGGPKATAA 291

Db 133 MYTDKKQVLSDDGMFLDYQVDTLEGSSGTVYDASHRVVGVHTLGDGANOINSVAKLNER 192

QY 292 FVEFI 296

Db 193 NLFFI 197

Search completed: June 21, 2002, 18:42:41

Job time: 125 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:42:21 ; Search time 13.5 Seconds  
(without alignments)  
866.171 Million cell updates/sec

Title: US-09-652-743A-12  
Perfect score: 1559  
Sequence: 1 MKKVKLIPSLVFGALSVP.....NGGPKATAAFVEFINAKAQ 302

ing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

earched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	420.5	27.0	316	1 GSEP_BACLI	P80057 bacillus li
2	280.5	18.0	313	1 MPR_BACSU	P39790 bacillus su
3	189.5	12.2	336	1 STSP_STAUA	P04188 staphylococ
4	153.5	10.0	280	1 ETA_STAUA	P09331 staphylococ
5	133	8.5	273	1 YDGD_ECOLI	P76176 escherichia
6	123	8.5	277	1 ETR_STAUA	P09332 staphylococ
7	123	7.9	1656	1 OMPB_RICJA	O06653 r outer mem
8	120	7.7	254	1 TRY3_AEDAE	P29786 aedes aegypt
9	119	7.6	678	1 YF48_MYCTU	Q10778 mycobacteri
10	118.5	7.6	227	1 TRYP_SACER	P24664 saccharopol
11	118	7.6	248	1 TRYP_FUSOX	P35049 fusarium ox
12	116.5	7.5	389	1 SERI_BOMMO	P07856 bombyx mori
13	116	7.4	280	1 TRY2_DROME	P42280 drosophila
14	113.5	7.3	256	1 HYPA_HYPLI	P35587 hypoderma l
15	113	7.2	569	1 YE59_MYCPN	P5324 mycoplasma
16	112	7.2	281	1 TRY2_DROER	P54630 drosophila
17	112	7.2	679	1 PAN3_YEAST	P36102 saccharomyc
18	111	7.1	1286	1 AIDA_ECOLI	Q03155 escherichia
19	110	7.1	253	1 CAC3_BOVIN	P05805 bos taurus
20	109	7.0	264	1 GRAK_HUMAN	P49863 homo sapien
21	109	7.0	532	1 SPG7_DICDI	P22698 dictyosteli
22	109	7.0	762	1 SLAP_ACEKI	P22258 acetogenium
23	108.5	7.0	1300	1 L20K_RICRI	P4914 rickettsia
24	108.5	7.0	1654	1 OMPB_RICPI	Q53047 r outer mem
25	108	6.9	262	1 TRY1_DROME	P42278 drosophila
26	108	6.9	1169	1 YR82_YEAST	P36170 saccharomyc
27	107.5	6.9	214	1 ACH2_LONAC	P23605 lonomia ach
28	107.5	6.9	348	1 ASG2_ECOLI	P00805 escherichia
29	107.5	6.9	1250	1 YFAL_ECOLI	P45508 escherichia
30	107	6.9	636	1 YNR6_YEAST	P33882 saccharomyc
31	106.5	6.8	497	1 FLIC_ECOLI	P04949 escherichia
32	106	6.8	664	1 Y4FB_RHISN	P55440 rhizobium s
33	106	6.8	2249	1 OMPA_RICRI	P15921 rickettsia

ALIGNMENTS

RESULT	ID	Score	Query Match %	Length	ID	Description
1	GSEP_BACLI	420.5	27.0	316	1 GSEP_BACLI	P80057 bacillus li
2	MPR_BACSU	280.5	18.0	313	1 MPR_BACSU	P39790 bacillus su
3	STSP_STAUA	189.5	12.2	336	1 STSP_STAUA	P04188 staphylococ
4	ETA_STAUA	153.5	10.0	280	1 ETA_STAUA	P09331 staphylococ
5	YDGD_ECOLI	133	8.5	273	1 YDGD_ECOLI	P76176 escherichia
6	ETR_STAUA	123	8.5	277	1 ETR_STAUA	P09332 staphylococ
7	OMPB_RICJA	123	7.9	1656	1 OMPB_RICJA	O06653 r outer mem
8	TRY3_AEDAE	120	7.7	254	1 TRY3_AEDAE	P29786 aedes aegypt
9	YF48_MYCTU	119	7.6	678	1 YF48_MYCTU	Q10778 mycobacteri
10	TRYP_SACER	118.5	7.6	227	1 TRYP_SACER	P24664 saccharopol
11	TRYP_FUSOX	118	7.6	248	1 TRYP_FUSOX	P35049 fusarium ox
12	SERI_BOMMO	116.5	7.5	389	1 SERI_BOMMO	P07856 bombyx mori
13	TRY2_DROME	116	7.4	280	1 TRY2_DROME	P42280 drosophila
14	HYPA_HYPLI	113.5	7.3	256	1 HYPA_HYPLI	P35587 hypoderma l
15	YE59_MYCPN	113	7.2	569	1 YE59_MYCPN	P5324 mycoplasma
16	TRY2_DROER	112	7.2	281	1 TRY2_DROER	P54630 drosophila
17	PAN3_YEAST	112	7.2	679	1 PAN3_YEAST	P36102 saccharomyc
18	AIDA_ECOLI	111	7.1	1286	1 AIDA_ECOLI	Q03155 escherichia
19	CAC3_BOVIN	110	7.1	253	1 CAC3_BOVIN	P05805 bos taurus
20	GRAK_HUMAN	109	7.0	264	1 GRAK_HUMAN	P49863 homo sapien
21	SPG7_DICDI	109	7.0	532	1 SPG7_DICDI	P22698 dictyosteli
22	SLAP_ACEKI	109	7.0	762	1 SLAP_ACEKI	P22258 acetogenium
23	L20K_RICRI	108.5	7.0	1300	1 L20K_RICRI	P4914 rickettsia
24	OMPB_RICPI	108.5	7.0	1654	1 OMPB_RICPI	Q53047 r outer mem
25	TRY1_DROME	108	6.9	262	1 TRY1_DROME	P42278 drosophila
26	YR82_YEAST	108	6.9	1169	1 YR82_YEAST	P36170 saccharomyc
27	ACH2_LONAC	107.5	6.9	214	1 ACH2_LONAC	P23605 lonomia ach
28	ASG2_ECOLI	107.5	6.9	348	1 ASG2_ECOLI	P00805 escherichia
29	YFAL_ECOLI	107.5	6.9	1250	1 YFAL_ECOLI	P45508 escherichia
30	YNR6_YEAST	107	6.9	636	1 YNR6_YEAST	P33882 saccharomyc
31	FLIC_ECOLI	106.5	6.8	497	1 FLIC_ECOLI	P04949 escherichia
32	Y4FB_RHISN	106	6.8	664	1 Y4FB_RHISN	P55440 rhizobium s
33	OMPA_RICRI	106	6.8	2249	1 OMPA_RICRI	P15921 rickettsia

RESULT 1  
GSEP\_BACLI  
ID GSEP\_BACLI  
AC P80057  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glutamyli endopeptidase precursor (EC 3.4.21.19) (Glutamate specific endopeptidase) (GSE).  
DE BLASE.  
GN Bacillus licheniformis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1402;  
RN [1]  
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=ATCC 14580;  
RX MEDLINE=93054737; PubMed=1429718;  
RA Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E., Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;  
RT \*Purification, characterization, cloning, and expression of a glutamic acid-specific protease from Bacillus licheniformis ATCC 14580.\*;  
RT J. Biol. Chem. 267:23782-23788(1992).  
[2]  
SEQUENCE OF 95-316.  
MEDLINE=92155199; PubMed=1346764;  
RX Svendsen I., Bredam K.;  
RT \*Isolation and amino acid sequence of a glutamic acid specific endopeptidase from Bacillus licheniformis.\*;  
RT Eur. J. Biochem. 204:165-171(1992).  
CC -|- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG PREFERENCE FOR GLU.  
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Asp-I-Xaa, Glu-I-Xaa.  
CC -|- SUBCELLULAR LOCATION: Secreted  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE V8 FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: D10060; BAA00949.1; -;  
DR PIR: S23078; S23078.  
DR PIR: A45134; A45134.  
DR MEROPS: S01.271; -;  
DR InterPro: IPR000126; Ser\_proteas\_v8.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: SM00839; V8PROTEASE.  
DR SMART: SM00020; Tryp\_Spc; 1.



```
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT PROPEP 31 94
FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 126 142
FT DISULFID 275 279
SQ SEQUENCE 316 AA; 33611 MW; 9607552CB7089B09 CRC64;

Query Match 27.0%; Score 420.5; DB 1; Length 316;
Best Local Similarity 33.0%; Pred. No. 4.1e-23;
Matches 101; Conservative 55; Mismatches 121; Indels 29; Gaps 10;

QY 5 KKLIPSLVPGALSPVFA--HAASDSVLTSYDMVTSDG--KVISSADFHNDKMT---56
R 4 KRSVKRLITGLIGISYLGMPA--QAAPSPHTPVSSDPSYKAETSVTYDNIKSDQY 61

b 57 ---PSFDFKDDLSSTIGERKVKPLTYLKDFTQKVVIGDDGRTKVTNTRVAPYNSIAYIT 113
62 GLYSKAFPTGKYNET--KEAEKSPAKAPYSIKSVIGSDRTRVNTTAYPYRAIVHIS 120

QY 114 FGSSCTGTLLIANKILTNHCHVYNTATRSYSAKGSYVPGMNDSTAVNGSANTFEYVPS 173
121 SSIGSCTGWMIGPKTVATAGHCYDTSSTGSGAGTATVSPGRNGTSYPYGSVKSTRYFIPS 180

QY 174 GYINTGASQYDFAVIKTDTNIGNTVGYRSIROVT-NLTGTTIKISYPGDKMRSTCKVSO 232
181 GW-RSGNTNDYGAIESEPIGNTVGTGYSYTTSLVGLITVITISGIPGDKTAGT---Q 235

QY 233 WEMSGPVTRDTNLYITIDTFSGNSGSAMLDQNOQ-----IVGVHNAG-YSNGTIN 283
236 WOHSGPIAISETKLYQYAMDYTGQSGSPVFEQSSSRNCSGPGSLAVHTNGYVGSSYN 295

QY 284 GGPAT 289
296 RGRIT 301

RESULT 2
MPR_BACSU STANDARD; PRT; 313 AA.
AC P39790;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Extracellular metalloprotease precursor (EC:3.4.21.-).
GN MPR.
Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
NCBI_TaxID=1423;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-GP241;
RX MEDLINE-90130256; PubMed-2105291;
RA Sloma A., Rudolph C.F., Rufo G.A. Jr., Sullivan B.J., Theriault K.A.,
RA Ally D., Pero J.;
RT "Gene encoding a novel extracellular metalloprotease in Bacillus
subtilis."
RL J. Bacteriol. 172:1024-1029(1990).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-168;
RA Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the
Bacillus subtilis chromosome."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 1-68 FROM N.A.
```

```
RX MEDLINE-89108019; PubMed-3145906;
RA Smith H., de Jong A., Bron S., Venema G.;
RT "Characterization of signal-sequence-coding regions selected from the
Bacillus subtilis chromosome."
RL Gene 70:351-361(1988).
RN [4]
RP SEQUENCE OF 1-10 FROM N.A.
RC STRAIN-168;
RX MEDLINE-96118702; PubMed-7496533;
RA Saxild H.H., Jacobsen J.H., Nygaard P.;
RT "Functional analysis of the Bacillus subtilis purt gene encoding
formate-dependent glycinamide ribonucleotide transformylase."
RL Microbiology 141:2211-2218(1995).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
V8 FAMILY.
CC -1- CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2
FAMILY OF SERINE PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L10505; AAA22604.1;
DR EMBL; AB006424; BAA33121.1;
DR EMBL; M22916; AAA22832.1;
DR EMBL; Z99105; CAB12018.1;
DR PIR; A35122; A35122.
DR HSP; P00756; ISGF.
DR MEROPS; S01.272;
DR Subtilist; BG10690; mpr.
DR InterPro; IPR000126; Ser.proteas_V8.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolase; Serine protease; Signal; Zymogen; Complete proteome.
FT SIGNAL 1 34 POTENTIAL.
FT PROPEP 35 93
FT CHAIN 94 313 EXTRACELLULAR METALLOPROTEASE.
FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 267 267 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 131 147 BY SIMILARITY.
FT CONFLICT 61 68 QVSAPYEG->PLESTAQA (IN REF. 3).
SQ SEQUENCE 313 AA; 33842 MW; D41788E8D652AE94 CRC64;

Query Match 18.0%; Score 280.5; DB 1; Length 313;
Best Local Similarity 29.4%; Pred. No. 3.6e-13;
Matches 96; Conservative 50; Mismatches 129; Indels 51; Gaps 15;

QY 6 KLIPSLV--FGALSPVFAHAASDS-----VLTSDYDMVTSDGKVISSADFHNDKMT 57
Db 2 KLIPRFKQWFAYLTVLCLALAAAVSFGVPAKAAENPQTSVSTGK---EAD-----ATK 53
QY 58 SSFDKVDLISST---IGERKVKPL---TYLKDFTQ---KVVIGDGRKVTNTRVAPYN 107
Db 54 NOTSKADQVSAPYEGTCKTSKSLYGGQTELEKNIOFPSSIIQTGTRISSTTFFPYR 113
QY 108 -----SIAY-----ITFGSSCTGTLLIANKILTNHCHVYNTATRSYSAKGSVYPGNMDS 158
Db 114 ATVOLSIKPYNTSSTYG---CTGLVNPNTVITAGHCYV-SQDHGWASTITAAPGRNGSS 169
QY 159 AVNGSANTFEYVPSGYINTGASQYDFAVIKTDTNIGNTVGYRSIROVTNLT--GTTIKI 216
Db 170 YPYGTVSGTMFYSVKGWTESKDTNIDYGAIKLNGSPCNTVGYRTNNTSSSPVGLSSV 229
```

QY 217 SGYPGDKMRSTCKVKSOWEMSGPVYREDTNTLAYTIDTFSNGSAML-----DONQOIVGV 272  
Db 230 TGPCKDKTGT-----MMSDKPIRSAETKLTYYTDTYGCQSGPVYRNSDTGQTAIAI 285  
QY 273 HNAGSYNGTINGKPKATAAFVEFINY 298  
Db 286 HTNGGS--SYNLGTRVNDVFNNOY 309

RESULT 3  
ID STSP\_STAAU STANDARD; PRT; 336 AA.  
AC P04188:  
DT 20-MAR-1987 (Rel. 04, Created)  
01-APR-1988 (Rel. 07, Last sequence update)  
01-MAR-2002 (Rel. 41, Last annotation update)  
Glutamyl endopeptidase precursor (EC 3.4.21.19) (Staphylococcal serine  
proteinase) (V8 proteinase) (Endoproteinase Glu-C).  
SSPA.  
Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-V8;  
RX MEDLINE-87316953; PubMed-3306605;  
RA Carmona C., Gray G.L.;  
RT "Nucleotide sequence of the serine protease gene of Staphylococcus  
aureus, strain V8."  
RL Nucleic Acids Res. 15:6757-6757(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20569178; PubMed-11119502;  
RA Rice K., Peralta R., Bast D., de Azavedo J., McGavin M.J.;  
RT "Description of staphylococcus serine protease (ssp) operon in  
Staphylococcus aureus and nonpolar inactivation of sspA-encoded serine  
protease."  
RL Infect. Immun. 69:159-169(2001).  
RN [3]  
RP SEQUENCE OF 69-280.  
RX STRAIN-V8;  
RX MEDLINE-78212487; PubMed-969922;  
RA Drapeau G.R.;  
RT "The primary structure of staphylococcal protease."  
RL Can. J. Biochem. 56:534-544(1978).  
-1- FUNCTION: PREFERENTIALLY CLEAVES PEPTIDE BONDS ON THE CARBOXYL-  
TERMINAL SIDE OF ASPARTATE AND GLUTAMATE.  
-1- CATALYTIC ACTIVITY: Preferential cleavage: Asp-I-Xaa, Glu-I-Xaa.  
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE  
V8 FAMILY.  
-1- DATABASE: NAME=Worthington enzyme manual;  
WWW="http://www.worthington-biochem.com/manual/P/STAP.html".  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
CC EMBL; Y00356; CAA68434.1; .  
CC EMBL; AF309515; AAG45843.1; .  
CC PIR; A26812; PRSASK.  
CC MEROPS; S01.269; .  
CC InterPro; IPR000126; Ser\_proteas\_v8.  
CC InterPro; IPR001254; Trypsin.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00839; V8PROTEASE.  
CC PROSITE; PS00672; V8\_HIS; 1.  
CC PROSITE; PS00673; V8\_SER; 1.  
CC Hydrolase; Serine protease; Zymogen; Signal.

FT SIGNAL 1 29  
FT PROPEP 30 68  
FT CHAIN 69 336  
FT ACT\_SITE 119 119  
FT ACT\_SITE 161 161  
FT ACT\_SITE 237 237  
FT CONFLICT 109 109  
FT CONFLICT 125 125  
FT CONFLICT 145 145  
FT CONFLICT 193 193  
FT CONFLICT 229 229  
FT CONFLICT 259 261  
FT CONFLICT 268 270  
FT SEQUENCE 336 AA; 36326 MW; 8B138D0C7996AA3E CRC64;  
POTENTIAL.  
GLUTAMYL ENDOPEPTIDASE.  
CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.  
MISSING (IN REF. 3).  
N -> D (IN REF. 3).  
V -> T (IN REF. 3).  
D -> N (IN REF. 3).  
EFN -> OFD (IN REF. 3).  
ENV -> NEYN (IN REF. 3).  
Query Match 12.2%; Score 189.5; DB 1; Length 336;  
Best Local Similarity 24.2%; Pred. No. 1.2e-06;  
Matches 74; Conservative 45; Mismatches 126; Indels 61; Gaps 12;  
QY 3 KYKKLIPSLLVFGALSVPSFAHAASDVLTSDYDMVTSQKVISSADFHNDMKTSPSFDK 62  
Db 7 KVSSLFVATLTATLVSSPAANALS-----SKAMDNHPQQTQSSKQ 48  
QY 63 VDDLSTIGEKKVPLTLYLKDFQTKVIGDDGRTKVTNTRVAPYNSIAYITF-----GGSS 118  
Db 49 TPRIQK--GGNLKPLE---QREHANYILPNDRHQITDTTNGHYAPVYTIQVEAPTGTFI 103  
QY 119 CTCTLIAPNKILTNGHCYVNTATRSYSAKSVYV-CMNDSTAVSGSANMTEFVVPVSGYIN 177  
Db 104 ASGVVGVKDTLLTNKHVVDATHGDPHALK--APPSAINQDNPNGGFTAEQITKYSG--- 158  
QY 178 TGASQYDFAVIKTDTNIGN-----TVGYRSIRSQVTNLTGTTIKISGYPGDKMRST 227  
Db 159 ----ECGLAIVKFSPEONKHGCEVVKPATMSNAETQVNO-----NITVTGYPCDKPVAT 210  
QY 228 GKVSQWEMSGPVYREDTNTLAYTIDTFSNGSAMLQOQOIVGVNAGYSNCTINGGPK 287  
Db 211 ----MWESKGRITLYLKGEMHQYDLSTTGCNCGSPVFNKNEKVIHGWGVPN-EFNG--- 262  
QY 288 ATAAFY 293  
Db 263 --AVFI 266

RESULT 4  
ETA\_STAAU STANDARD; PRT; 280 AA.  
AC P09331;  
DT 01-MAR-1989 (Rel. 10, Created)  
01-MAR-1989 (Rel. 10, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Exfoliative toxin A precursor (EC 3.4.21.-) (Epidermolytic toxin A).  
GN ETA.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-UT0002;  
RX MEDLINE-87307959; PubMed-3040666;  
RA Lee C.Y., Schmidt J.J., Johnson-Winegar A.D., Spero L., Iandolo J.J.;  
RT "Sequence determination and comparison of the exfoliative toxin A and  
toxin B genes from Staphylococcus aureus."  
RL J. Bacteriol. 169:3904-3909(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TC16;  
RX MEDLINE-87307960; PubMed-3040667;  
RA O'Toole P.W., Foster T.J.;  
RT "Nucleotide sequence of the epidermolytic toxin A gene of  
Staphylococcus aureus."

pfam: PF00089; trypsin: 1.  
PRINTS: PR00839; V8PROTEASE.  
PROSITE: PS00672; V8\_HIS: 1.  
PROSITE: PS00673; V8\_SER: 1.  
KWK Hydrolyase; Serine protease; Calcium; Toxin; Signal; 3D-structure.  
FT SIGNAL 1 38  
CHAIN 39 280 EXFOLIATIVE TOXIN A.  
FT ACT\_SITE 110 110 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 158 158 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 233 233 CHARGE RELAY SYSTEM.  
FT MUTAGEN 233 233 S->G: LOSS OF TOXICITY.  
SEQUENCE 280 AA; 31077 MW; 4BFA750FFE43586 CRC64;

Query Match 10.0%; Score 155.5; DB 1; Length 280;  
Best Local Similarity 24.7%; Pred. No. 0.00025;  
Matches 77; Conservative 46; Mismatches 132; Indels 57; Caps 13;

OY 2 KVKKKLIPSLVFGALSVPSFAHAASDVLTSDYDMV-TSDKGVVISSADFHNMDKTPSSF 60  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 5 KIISKVLLLSLF---TVGSAFEIVODELMOKNHAKAEVS AEIKHKEEKNNYYGYNAF 61  
  
OY 61 DKVDLDSITIGEEKVKPLTTYLKDFQTKVVIGDDGRTKVTRVPAPNSIAYI-TFGGSSC 119  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 62 NLPKELFSKVEK-----DRQKYPTNTIGNVFVKQTSA 95  
  
OY 120 TGTLIAPNKLITNGHCYINTATRSYSAKSQSYVPCMDNSTAVNGSANMTEFYVPSGYINTG 179  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 96 TGVVLCKNTVLTNRHIA--KFANGDP SKVSFRPSINTDDNGTETPYGEYEVRKILQEPF 153  
  
OY 180 ASOYDPAVI--KTDTN---IGNTVGYRSTRQVTNL-TGTTIKISGYPGDKMIRSTGKYSOW 233  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 154 GAGVDIALIRLPDQNGVSLGDGISPAKITGSNDLKDGDKLEILIGYPFDH-----KYNQM 208  
  
OY 234 ENSG-PVTREDTNLAYTIIDTFSGNSGAMLDNQOOIVGVHNAAGYSN----GTINGPKA 288  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 209 HNSEIELTWLRGLRYGP-TVPCNGSGGIFNSGELVGIHSSKVSCHLDREHQINYC--- 264  
  
OY 289 TAAFVEFINYAK 300  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 265 ----VGIGNYVK 272

RESULT 5  
YDGD\_ECOLI STANDARD; PRT; 273 AA.

ID YDGD\_ECOLI AC P76176:  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE DE Putative protease ydgd precursor (BC 3.4.21.-).  
GN YDGD OR B1598.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX Escherichia.  
NCBI\_Taxid=562;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.C., Rose D.J.,  
RA Mau B., Shao Y.  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
CC -I SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE  
CC V8 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL: AE000255; AAC74670.1;  
DR MEROPS: S01.260;  
DR Ecogen: EGI13925; ydgD.  
DR InterPro: IPR000126; Ser\_proteas\_v8.  
DR InterPro: IPR001254; Trypsin.  
DR PRINTS: PR00839; V8PROTEASE.  
DR SMART: SM00020; TRYP\_SPC.1.  
DR PROSITE: PS00134; TRYP\_SIN\_HIS.1.  
DR PROSITE: PS00135; TRYP\_SIN\_SER.1.  
DR Hypothetical protein: Hydrolase; Serine protease; Signal;  
KW Complete proteome.  
CC SIGNAL 1 21 POTENTIAL.  
CC CHAIN 22 273 PUTATIVE PROTEASE YDGD.  
CC ACT\_SITE 84 84 CHARGE RELAY SYSTEM (BY SIMILARITY).  
CC ACT\_SITE 223 223 CHARGE RELAY SYSTEM (BY SIMILARITY).  
SQ SEQUENCE 273 AA; 29277 MW; C8FCD018A59DBC62 CRC64;

Query Match 9.8%; Score 153.5; DB 1; Length 273;  
Best Local Similarity 24.6%; Pred. No. 0.00033;  
Matches 55; Conservative 37; Mismatches 89; Indels 43; Gaps 9;  
QY 83 DFQTKVIGDGRKTKVTRVAPYNSIAYI-TFGSSCTGTLIAPNKLITNGHCVYN---138  
DB 32 DEVSTLFFGHDDRPVNDTQSPWDVAGQLETASGNLCTATLIAPNLTAGHCLLTTPK 91  
QY 139 -----TATRSYSAK-----GSVYPMNDSTAVNGSANMTEFYVPSGYNLTGAS 181  
DB 92 GRADRAVALRFSNKGMLWRYEIHDEGRVDPITLGRKLKADGDC---WIVP-----PAAA 142  
QY 182 QYDFA-VIKTDNIGNTV-----GYRSIQVNTLTGTTIKISGYPGDKMRSTGRKVSQWE 234  
DB 143 PMDFGLVILRNPPSGITPLPLEFGDKAALTAALKAGRKVTQAGYVEDHLDLTLYSHQNC 202  
QY 235 MGPVTRDTNLAYITIDTFSGNSGAML---DONQOIVGVHNA 275  
DB 203 VTG---WAQTSVMSHOCDLPGDSGSLMLHTDDGWLIGVQSS 243

RESULT 6  
ETB\_STAATU STANDARD; PRT; 277 AA.  
P09332:  
01-MAR-1989 (Rel. 10, Created)  
01-MAR-1989 (Rel. 10, Last sequence update)  
01-NOV-1995 (Rel. 32, Last annotation update)  
DE Exfoliative toxin B precursor (EC 3.4.21.-) (Epidermolytic toxin B).  
GN ETB.  
OS Staphylococcus aureus.  
OG Plasmid pRW001.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86277939; PubMed=3733674;  
RA Jackson M.P., Iandolo J.J.;  
RT "Sequence of the exfoliative toxin B gene of Staphylococcus aureus.";  
RL J. Bacteriol. 167:726-728(1986).  
RN [2]  
RP REVISIONS, AND PARTIAL SEQUENCE.  
RC STRAIN=UT0002.  
RX MEDLINE=87307959; PubMed=3040666;  
RA Lee C.Y., Schmidt J.J., Johnson-Winegar A.D., Spero L., Iandolo J.J.;  
RT "Sequence determination and comparison of the exfoliative toxin A and  
RT toxin B genes from Staphylococcus aureus.";  
RL J. Bacteriol. 169:3904-3909(1987).  
RN [3]  
RP FUNCTION.

RX MEDLINE=90351379; PubMed=2117445;  
RA Bailey C.J., Smith T.P.;  
RT "The reactive serine residue of epidermolytic toxin A.";  
RL Biochem. J. 269:535-537(1990).  
RN [4]

RP FUNCTION.  
RX MEDLINE=90346117; PubMed=2384148;  
RA Dancer S.J., Garrat R., Saldanha J., Jhoti H., Evans R.;  
RT "The epidermolytic toxins are serine proteases.";  
RL FEBS Lett. 268:129-132(1990).

CC -!- FUNCTION: HAS SERINE PROTEASE-LIKE PROPERTIES AND BINDS TO THE  
CC SKIN PROTEIN PROFILAGGRIN  
CC -!- DISEASE: EXFOLIATIVE TOXINS CAUSE IMPETIGOUS DISEASES COMMONLY  
CC REFERRED AS STAPHYLOCCAL SCALDED SKIN SYNDROME (SSSS).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE  
CC V8 FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL: M17348; AAA26628.1;  
DR PIR: B26680; PRSAEB.  
DR PIR: A26050; A26050.  
DR HSP: P09331; IAGJ.  
DR MEROPS: S01.270;  
DR InterPro: IPR000126; Ser\_proteas\_v8.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00839; V8PROTEASE.  
DR PROSITE: PS00672; V8\_HIS.1.  
DR PROSITE: PS00673; V8\_SER.1.  
DR Hydrolase; Serine protease; Toxin; Signal; Plasmid.  
KW SIGNAL 1 31  
FT CHAIN 32 277 EXFOLIATIVE TOXIN B.  
FT ACT\_SITE 96 96 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 145 145 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).  
SQ SEQUENCE 277 AA; 30765 MW; DE1A7BEE273CF92F CRC64;

Query Match 8.5%; Score 133; DB 1; Length 277;  
Best Local Similarity 28.7%; Pred. No. 0.0097;  
Matches 56; Conservative 26; Mismatches 79; Indels 34; Gaps 9;

QY 96 TKVTNTRVAPYNSIAYITFGSS-CTGTLIAPNKLITNGHCVYNTATRSYSAKSYVPGM 154  
DB 57 THITDNARSPYNSVGVTFVKGSTLATGVLGKNTIVTNVH-VAREAAKNPS-----NI 108  
QY 155 NDSTAVNGSANMTEFYVPSGYINT---GASQY-----DFAVIKTDN-IGNTVGYRSIQV 206  
DB 109 IFTPAQRDAEKNEFPTPYGKFAEEIKESPGQGLDLAIILKPKKEKESAG--DLIQP 166  
QY 207 TNL-----TGTTIKISGPGDKMRSTCKVSWENSGPVTREDTNLAYITIDTFSGNSG 259  
DB 167 ANIPDHIDIOKGDYSLGYPYNSAYSILYQSIEM-----FNDSQYFGYTEVNGSG 218  
QY 260 SAHLNQNOQIVGVHN 274

DB 219 SGIFNLKGLIGHS 233  
RESULT 7  
OMP\_RICJA  
ID OMP\_RICJA STANDARD; PRT; 1656 AA.  
AC O06653;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
DE (rOmpB protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)  
DE [Contains: 120 kDa surface-exposed protein (Surface protein  
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
GN OMPB  
OS Rickettsia japonica.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=35790;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-YH;  
RA Uchiyama T.;  
RT "sequencing of the gene encoding the protein rOmp B of Rickettsia  
RT japonica";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
CC SIMILARITY).  
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AB003681; BAA20138.1;  
CC InterPro: IPR003858; OmpA\_rOmpB.  
CC Pfam: PF02708; rOmpA\_rOmpB; 1.  
KW Antigen; S-layer; Cell wall.  
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.  
FT DOMAIN 528 533 POLY-GLY.  
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;  
Query Match 7.9%; Score 123; DB 1; Length 1656;  
Best Local Similarity 23.3%; Pred. No. 0.44;  
Matches 90; Conservative 52; Mismatches 105; Indels 140; Gaps 25;  
QY 4 VKKLIPSLVFG--ALSVPSFAHRAASDVL-----TSYDMVTSYG-----42  
Db 8 LKKLISAGLVTAATVAFSAGSAMGAATQONRTTNGVATTVDGVGFDQTVALANVAYA 67  
43 -KVISSADFNH--DMKTPS-SFD-----KYDDLSSSTIGK-----VKPLTYLTKDFOTKV 88  
68 PNAVITANANNINGINLNPAGSFNGFLSNANNLAVTSEDTTLCGFINNANNANRFLNLT 127  
89 VIGDDGRT-----KVTNTR-----VAPYN-----SIAYITGGSSCT- 120  
128 ---DAGKTLTTGGITNVQSAATHNAQNIIVAKFNGCGAATAMNDLSGLGTTIDFGAAASTL 184  
121 -----GTLIAPNKILNGHCYVYNTATRSYSKAGSVYPGMNDSTAVNGSANMTE 168  
185 VFDLANPTTOKAPLILADNALIVNG-----ANGTLNVT- 217  
169 FYVPSGYINTGASQYDFAVIKTDNIGNITVGYRSIROVTNLT-----GTTIKISGYPG 221  
218 ---NGFIQ--VSDKSFATVKA--INIGDGGCFMNTNANNALNLOAGGTTINFNGTDG 270  
222 DKMRSTGKVSQWEMSGPVTNLTAYITDITFSGN--SGSAMLQD-----NQIIVGVHNAGY 277  
271 -----TGRVLLSLKNGAAT--DFNVT-----GSLGSLNKGIIELNTVAINGOLIA--NAGP 317  
QY 278 SN-----GTINGGPKATAAFVEFINTAKA 301

Db 318 ANAVICTNNGAGRA-AGFVVSVVDNGKA 343  
RESULT 8  
TRY3\_AEDAE  
ID TRY3\_AEDAE STANDARD; PRT; 254 AA.  
AC P29786;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Trypsin 3AL precursor (EC 3.4.21.4).  
OS Aedes aegypti (Yellowfever mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
OC Aedes.  
OX NCBI\_TaxID=71159;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97242546; PubMed=9087545;  
RA Kalhok S., Tabak L.M., Prosser D.E., Brook W., Downer A.E.R.,  
RA White B.N.;  
RT "Isolation, sequencing and characterization of two cDNA clones coding  
RT for trypsin-like enzymes from the midgut of Aedes aegypti";  
RL Insect Mol. Biol. 2:71-79(1993).  
CC -!- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD  
CC MEAL.  
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- TISSUE SPECIFICITY: MIDGUT.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X64362; CAA45714.1;  
CC FIR: S19890; TRWV3Y.  
CC HSSP: P00763; LDPO.  
CC MEROPS: S01.112;  
CC InterPro: IPR001314; Chymotrypsin.  
CC InterPro: IPR001254; Trypsin.  
CC Pfam: PF00089; trypsin; 1.  
CC PRINTS: PR00722; CHYMOTRYPSIN.  
CC SMART: SM00020; Tryp\_SPC; 1.  
CC PROSITE: PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE: PS00134; TRYPSIN\_HIS; FALSE\_NEG.  
CC PROSITE: PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Signal; Zymogen; Digestion;  
KW Multigene family.  
FT SIGNAL 1 20 POTENTIAL.  
FT PROPEP 21 27 ACTIVATION PEPTIDE.  
FT CHAIN 28 254 TRYPSIN 3AL.  
FT ACT\_SITE 68 68 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 113 113 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 209 209 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 53 69 BY SIMILARITY.  
FT DISULFID 178 194 BY SIMILARITY.  
FT DISULFID 205 229 BY SIMILARITY.  
FT SITE 203 203 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
SQ SEQUENCE 254 AA; 26939 MW; 43AAF13B8DE7D485 CRC64;  
Query Match 7.7%; Score 120; DB 1; Length 254;  
Best Local Similarity 25.5%; Pred. No. 0.073;  
Matches 48; Conservative 31; Mismatches 73; Indels 36; Gaps 10;  
QY 115 GGSCTGTLIAPNKILNGHCYVYNTATRSYSKAGSVYPGMNDSTAVNGSANM-TEFYVPS 173

```
Db 49 GRHFCGSIISPRWVLTTRAHCTNTDPAAYTIRA-----GSTDRT--NGGIIVKVKSVIPH 102
QY 174 GYINTGASQYDFAVIKTNTNIGNTVGYRQV-----TNLTGTTIKISGYPGDK--- 223
Db 103 POYNGDYNDFSLLEDESIGFS---RSIEALPDASETVADGAMCTVSGWGTKNVF 159
QY 224 -----MRSTG--KVSQWMSG-----PVTRDTNLAAYT--IDTFSGNSGSAMLDQNOO 268
Db 160 ENMTLLRAVNPVPSYNOACAAALVNVVPTQMICAGYAAGKDKSCQSDSGPLV-SGDK 218
QY 269 IVGVHINAG 276
Db 219 LVGVWSMG 226

LT 9
MYCTU
YF48_MYCTU STANDARD; PRT; 678 AA.
Q10778;
AC 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV1548c.
GN RV1548C OR MT1599 OR MTCr48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feitwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT *Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.*;
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RX Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RX Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RX Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RX Bishai W.;
RT *Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z74020; CA98335.1; -;
CC EMBL; AE007026; AAK45866.1; ALT_INIT.
CC TIGR; MT1599; -;
CC TubercuList; RV1548c; -;
CC InterPro; IPR002989; Mycobact_pentapep.
CC InterPro; IPR000030; PPE.
CC Pfam; PF01469; Pentapeptide_2; 13.
CC Pfam; PF00823; PPE; 1.
```

```
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT CONFLICT 258 D -> G (IN REF. 2).
SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;

Query Match 7.6%; Score 119; DB 1; Length 678;
Best Local Similarity 25.7%; Pred. No. 0.29;
Matches 80; Conservative 41; Mismatches 112; Indels 78; Gaps 18;

QY 14 FGALSVPS-----FAHAASDSV--LTSDDYDMV-----TSDGKVISSADFNDMKTTPSSFDKV 63
Db 379 FGALNSGSGNLGFGNSGNGNIGFFNSGNNIGMNSGNGVGLSVFSGSAERSGSGNS 438
QY 64 DDLSSITGEKVKVPLTTLTKDFQTKVVGIDGRTK-----VTNTRVAPYNSIAYITFGG 116
Db 439 GELSTGIGNSGQLSTGWFNSATTSTGWFNSGTTNTGWFNSGTTNTGIG--NS-----GG 490
QY 117 SSCTGTLIAPNKLITNGHCYVNTATRSYSAKGSVYPMNDSTAVNGSANMTFFYVPSGYI 176
Db 491 NLVTGSM-----GLFNSGH-----TNTGSFNA-GSM-----NTGDFNSGNNVTGYFNSGNI 535
QY 177 NTGASQYDFAVIKTDTNIG--NTVGYRSTRQVNTLTG-----TTIKISGYPGDKMRSTGKVS 231
Db 536 NTGF-----FNSGDLNTGLFNSVYNQVQNSGWLHTGTNNISGYANAGTFNSG--- 581
QY 232 QWEMSGPVTRDTNLAAYTIDTFSGNSGSAML-DONQOI--VGVNAGYSNG--TINGGP 286
Db 582 -----FDNNARDEHAEEFTGNSGLANVGNINAGIIVGDHLSGFRNSVPTITG--- 629
QY 287 KATAAFVEFIN 297
Db 630 --TANISGFVN 638

RESULT 10
ID TRYP_SACER STANDARD; PRT; 227 AA.
AC P24664;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Trypsin (EC 3.4.21.4) (SET).
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE.
RA Miyamoto K., Matsuo H., Narita K.;
RL (In) Dai 30 kai tanpakushitsu kouzou touronkai kouen youshishuu,
RL pp.77-80 (1979).
RN [2]
RP REVISIONS.
RX MEDLINE=92176171; PubMed=1794983;
RA Nagamine Y., Norioka S., Sakiyama F.;
RL Unpublished results, cited by:
RL Yamane T., Kobuke M., Tsutsui H., Toida T., Suzuki A., Ashida T.,
RL Kawata Y., Sakiyama F.;
RL J. Biochem. 110:945-950(1991).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=92176171; PubMed=1794983;
RA Yamane T., Kobuke M., Tsutsui H., Toida T., Suzuki A., Ashida T.,
RA Kawata Y., Sakiyama F.;
RT *Crystal structure of Streptomyces erythraeus trypsin at 2.7-A
RT resolution.*;
RL J. Biochem. 110:945-950(1991).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- MISCELLANEOUS: HARDLY AUTOLYZES ITSELF AT ALL AT ITS ACTIVE PH
CC RANGE.
```

CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPsin FAMILY.  
 CC HSP; P20231; IAAO.  
 CC MEROPS; S01.102; -  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPsin.  
 DR SMART; SM00020; TRyp\_SPC; 1.  
 DR PROSITE; PS0240; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KW Hydrolase; Serine protease.  
 FT ACT\_SITE 42 42 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 88 88 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 179 179 CHARGE RELAY SYSTEM.  
 FT DISULFID 27 43  
 FT DISULFID 150 164  
 FT DISULFID 175 199  
 FT SITE 173  
 SEQUENCE 227 AA; 23308 MW; D5AC5E47B227B418 CRC64;  
 Query Match 7.6%; Score 118.5; DB 1; Length 227;  
 Best Local Similarity 25.0%; Pred. No. 0.082;  
 Matches 54; Conservative 33; Mismatches 82; Indels 47; Gaps 12;  
 QY 88 VWIGDGRKVTNTRVAPYNSIAVITFGSS-CTGTLIAPNKILTNCHVCYNTATRSYSA 146  
 Db 1 IVGED-----ANVDHPF-TVALVTPDGOQFCGTLAAPNKVTAHCTVG---SOPA 50  
 QY 147 KGSYPGMN-DSTAVNGSANMTEFYVPSYINTGASDYFAVTKTDNI-----GN 196  
 Db 51 DINVSGRTVMNSNIGTVSKVTVVWVHPEY-QDAAGFVSVTLTLEAPVKEAPIELAKAD 109  
 QY 197 TVGYSIRQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRDT-NLAY--YT--- 250  
 Db 110 DAGYAPDTAATILGWNTSGGQADHLQ-----KATPVNSDDTKQAYGEXTTNA 161  
 QY 251 -----IDTFSGNSGSAMLDQNOOIVGVHNAG 276  
 Db 162 MVCAGVEGGVDTCQDGGSGPMV-VNNKLIGVTSWG 196  
 RESULT 11  
 TRYP\_FUSOX  
 ID TRYP\_FUSOX STANDARD; PRT; 248 AA.  
 AC P35049;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 Trypsin precursor (EC 3.4.21.4).  
 Fusarium oxysporum.  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreales; mitosporic Hypocreales; Fusarium.  
 NCBI\_TaxID=5507;  
 [1]  
 RN SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=93324511; PubMed=8332590;  
 RA Rypniewski W.R., Hastrup S., Betzel C., Dauter M., Dauter Z.,  
 Papendorf G., Branner S., Wilson K.S.;  
 RT "The sequence and X-ray structure of the trypsin from Fusarium  
 oxysporum";  
 RL Protein Eng. 6:341-348(1993).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.55 ANGSTROMS).  
 RA Rypniewski W.R., Dambmann C., von der Osten C., Dauter M.,  
 RA Wilson K.S.;  
 RT "Structure of inhibited trypsin from Fusarium oxysporum at 1.55 A.";  
 RL Acta Crystallogr. D 51:73-84(1995).  
 CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-xaa, Lys-I-xaa.  
 CC -|- SUBCELLULAR LOCATION: Secreted.  
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPsin FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 DR EMBL; S63827; AAB27568.1;  
 DR PDB; 1TRY; 01-JAN-96.  
 DR MEROPS; S01.103; -  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPsin.  
 DR SMART; SM00020; TRyp\_SPC; 1.  
 DR PROSITE; PS0240; TRYPsin\_DOM; 1.  
 DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KW Hydrolase; Serine protease; zymogen; Signal; 3D-structure.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT PROPEP 18 25 ACTIVATION PEPTIDE.  
 FT CHAIN 26 248 TRYPsin.  
 FT ACT\_SITE 65 65 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 108 108 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 204 204 CHARGE RELAY SYSTEM.  
 FT DISULFID 50 66  
 FT DISULFID 174 189  
 FT DISULFID 200 225  
 FT SITE 198 198 REQUIRED FOR SPECIFICITY.  
 FT SEQUENCE 248 AA; 24576 MW; 1A0EBA88C3E70294 CRC64;  
 Query Match 7.6%; Score 118; DB 1; Length 248;  
 Best Local Similarity 21.7%; Pred. No. 0.099;  
 Matches 46; Conservative 42; Mismatches 88; Indels 36; Gaps 9;  
 QY 115 GSSCTGTLIAPNKILTNCHVCYNTATRSYSAKGSVYPCGMNDSTAVNGSANMTEFYVPSG 174  
 Db 46 GGPWCGSLLNANTVLTAAHCVSAGYAGSGFOIRA---GSLRSTGGITSLSSVRVHPS 101  
 QY 175 YINTGASDYFAVTKTDNI--GNTVGYRSIRQVTN--LTGTTIKISGYPG-DKMRSTGK 229  
 Db 102 Y---SGNNDLAILKILSTISPGSGNIGYARLAASGSDPVAGSATVAGWGATSEGGSTP 158  
 QY 230 VSQWEMSGPVTRDTNLAAYTI-----DTFSGNSGSAMLDQNOOIVGV 272  
 Db 159 VNLKVTVPVSVRATCRATQIGTSAITNOMFCAGVSSGGKSDCGDGGPIVDSNTLIGA 218  
 QY 273 HNAGYSNG-----TINGPKATAAFVEFIN-YA 299  
 Db 219 --VSWGNGCARPNYSGVYASVGALRSFIDTYA 248  
 RESULT 12  
 SERI\_BOMMO  
 ID SERI\_BOMMO STANDARD; PRT; 389 AA.  
 AC P07856;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sericin precursor (Silk gum protein).  
 OS Bombyx mori (Silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Bombycoidea; Bombycidae; Bombyx.  
 NCBI\_TaxID=7091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Middle silk gland;  
 RX MEDLINE=83082839; PubMed=6294094;

```

Okamoto H., Ishikawa E., Suzuki Y.;
"Structural analysis of sericin genes. Homologies with fibroin gene
in the 5' flanking nucleotide sequences.";
J. Biol. Chem. 257:15192-15199(1982).
[2]
SEQUENCE OF 317-354 FROM N.A.
MEDLINE=87076763; PubMed=3024742;
Michaille J.J., Coublie P., Prudhomme J.-C., Garel A.;
"A single gene produces multiple sericin messenger RNAs in the silk
gland of Bombyx mori.";
Biochimie 68:1165-1173(1986).
CC -|- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.
CC ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)
SECTION OF SILK GLANDS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licensese@isb-sib.ch).
-----
EMBL; J01029; - ; NOT_ANNOTATED_CDS.
DR EMBL; J01030; - ; NOT_ANNOTATED_CDS.
EMBL; J01031; - ; NOT_ANNOTATED_CDS.
DR EMBL; J01032; - ; NOT_ANNOTATED_CDS.
EMBL; J01033; - ; NOT_ANNOTATED_CDS.
DR EMBL; J01034; - ; NOT_ANNOTATED_CDS.
EMBL; M26101; AAA27843.1; -.
DR EMBL; M26102; AAA27844.1; -.
KW Silk; Signal.
FT SIGNAL. 1 15 POTENTIAL.
FT CHAIN 16 389 SERICIN.
FF SEQUENCE 389 AA; 38820 MW; BA605BC0305EAF19 CRC64;
SQ
Query Match 7.5%; Score 116.5; DB 1; Length 389;
Best Local Similarity 20.7%; Pred.No. 0.22;
Matches 67; Conservative 48; Mismatches 138; Indels 71; Gaps 11;
QY 12 LVFGALSVSPFAHAASD---SVLTSDYMVTSDGKVIS-----SADFHDMKTSSFD 61
Db :|::|||::| | | ::|::|::|::|::|::|::|::|::|::|::|::|::|::|
9 LIALALSVAFGHPGCRDTVEVKRKNNAASSESYLNKDNDISAGAHRAKSVEQSOD 68
62 KYDDLSTIGEKKVPITTYLKDPQTQVVIGDDGRVTNTRV-----APYN 107
69 KSKYTSGPEGVSTSGRSQNYKDKSQAIL---SGTGKSSNSNVQSDERKSASQSSSRSSQE 125
108 SIATITFGGSCCTGLIAPHNKILTNCHGVNTATRYSAGKSVYPGMN----DSTA VNGS 163
126 SASYSRSSSSSTSESSSSSSAAASTDASNTDSNSAGSTGGRRRTGYGSNSRDGS 185
164 ANMTEFPVPSGYINTGASDYDFAVIKTDIGN-----TVGYRSIRQVTNLGTGT 213
186 VS-----STGSSS-----NTDSNSSNAGSSTSOGESTYGYSS-----NSRDGS 223
214 IKISYGPDKMRSTGKVSOWHEMGPVTRDETNIAYTTIDPFSGNSSGAMLQQNOQIVGVH 273
224 VSTTGSSSNWNTDSNSVGSRRGSGSSHEDSKS--RDEVNSTTGSSNTDSNSNSVSGSS 281
QY 274 NA-----CYSNCTINGGPKATAA 291
Db :|::|::|::|::|::|::|::|::|::|::|::|::|::|
282 TSGGRRRTGYSSNRDGSVSTSGS 305
RESULT 13
TRYZ_DROME ID TRYZ_DROME STANDARD; PRT; 280 AA.
AC P42280; AC
DT 01-NOV-1995 (Rel. 32, Created)
```

```

01-NOV-1995 (Rel. 32, Last sequence update)
01-OCT-1996 (Rel. 34, Last annotation update)
Trypsin zeta precursor (EC 3.4.21.4).
TRY-ZETA.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[11]
SEQUENCE FROM N.A.
STRAIN=OREGON-R.
Wang S., Magoulas C., Hickey D.A.:
Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
-!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.
-!- SUBCELLULAR LOCATION: Extracellular.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U04853; AAA17456.1; -
HSSP; P35049; 1TRY.
MEROPS; S01.112; -.
FlyBase; FBgn0011556; Try-zeta.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR001254; Trypsin.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00020; Tryp_Spc; 1.
PROSITE; PS50240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Hydrolase; Serine protease; zymogen; Signal; Multigene family.
SIGNAL 1 22 PROBABLE.
PROPEP 23 38 ACTIVATION PEPTIDE.
CHAIN 39 280 TRYPSIN ZETA.
ACT_SITE 87 87 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 234 234 CHARGE RELAY SYSTEM (BY SIMILARITY).
DISULFID 72 88 BY SIMILARITY.
DISULFID 198 218 BY SIMILARITY.
DISULFID 230 254 BY SIMILARITY.
SITE 228 228 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SEQUENCE 280 AA; 29598 MW; 6255FICCEAE823CB CRC64;
-----
Query Match 7.4%; Score 116; DB 1; Length 280;
Best Local Similarity 27.5%; Pred. No. 0.16;
Matches 63; Conservative 26; Mismatches 104; Indels 36; Gaps 14;

Oy 81 LKDFQTKVIGDGR- --KVNTNRVAPYN-STAY---ITFGGS-----SCTGTLIAPNKIL 130
Ddb 26 LEDLDEKSV--PDGRIVGVGYATDIAQVYQISLRYGKIGITPENPFRRHRCGGSIFNETTV 83
Oy 131 TNGHCYVNTATRSYSAKGSVYPCMDNSTAVNG-SANMTEFYVPVGINTGASQYDFAVIK 189
Ddb 84 TAGHCVIGTVASQY----KVGAGNTPQTSQDGVITNKEIVHHEGYISGAAYNNDIAILF 139
Oy 190 TDTNIG-NTVGYRISROVFN--LTGTITIKISG----PGDKMRSTGKVSQWMSGPVTR 242
Ddb 140 VDPPLALNFTIKIGIKLASEQPIEGTVSKVSGWGTTSFG-----GYSSNQLLAVDVPV 193
Oy 243 DFNLAYYYTIDTSGNS---GSAMLDQNOQIVGVHNAGYSNGTINGGPKA 288
Ddb 194 SNELCDQDYEDFQETRYITTSAMLCAGKP--GVGGADAGQGD--SGGPLA 239

```

**RESULT 13**

TRYZ\_DROME

ID	TRYZ_DROME	STANDARD;	PRT;	280 AA.
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				

AC P42280;  
DT 01-NOV-1995 (BOL 33 Created)

DATE: 3/27/2001 11:00 AM



```

RESULT 14
HYPA_HYPLI
ID HYPA_HYPLI STANDARD; PRT: 256 AA.
AC P35587: Q25080;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypodermin A precursor (EC 3.4.21.-) (HA).
OS Hypoderma lineatum (Early cattle grub) (Common cattle grub).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Oestroidea; Oestridae; Hypoderma.
OX NCBI_TaxID=7389;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95107346; PubMed=7808473;
RA Moire N., Bigot Y., Periquet G., Boulard C.;
RT "Sequencing and gene expression of hypodermins A, B, C in larval
  stages of Hypoderma lineatum.";
Mol. Biochem. Parasitol. 66:233-240(1994).
[2]
SEQUENCE OF 3-256 FROM N.A.
Kuhn I., Files J.G., Pruett J.H., Temeyer K.B.;
Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 31-51.
RX MEDLINE=81232922; PubMed=7018579;
RA Tong N.T., Imhoff J.M., Lecroisey A., Keil B.;
RT "Hypodermin A, a trypsin-like neutral proteinase from the insect
  Hypoderma lineatum.";
Biochim. Biophys. Acta 658:209-219(1981).
CC -1- FUNCTION: SPECIFICITY, LIMITED TO CARBOXYL SIDE OF ARGININE
  RESIDUE IN B-CHAIN OF INSULIN.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DEVELOPMENTAL STAGE: LARVAL-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X74303; CAA52356.1; -
EMBL; L24914; AAA29224.1; -
PIR; A21590; A21590.
HSP; P00763; IDPO.
MEROPS; S01.111; -.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR001254; Trypsin.
Pfam; PF00089; trypsin; 1.
PRINTS; P00722; CHYMOTRYPSIN.
SMART; SM00020; TRYP_SPC; 1.
PROSITE; PS0240; TRYPsin DOM; 1.
PROSITE; PS00134; TRYPsin HIS; 1.
PROSITE; PS00135; TRYPsin SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 30 ACTIVATION PEPTIDE.
FT CHAIN 31 256
FT ACT_SITE 71 71 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 56 72 BY SIMILARITY.
FT DISULFID 180 197 BY SIMILARITY.
FT DISULFID 206 230 BY SIMILARITY.
FT SITE 204 204 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
FT CONFLICT 37 37 S -> M (IN REF. 3).
FT CONFLICT 45 45 Q -> E (IN REF. 3).
-----
FT CONFLICT 51 51 D -> P (IN REF. 3).
FT CONFLICT 248 248 L -> W (IN REF. 2).
SQ SEQUENCE 256 AA; 27907 MW; 63D12D337F66AA61 CRC64;

Query Match 7.3%; Score 113.5; DB 1; Length 256;
Best Local Similarity 22.7%; Pred. No. 0.21;
Matches 55; Conservative 42; Mismatches 76; Indels 69; Gaps 14;

OY 66 LSSTIGEKYKPLTYLKDFQTKVVGDDGRTKVTNTRVAPYNSIAVITFGSSCTGTLLA 125
  || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 24 LSQSDGRIVGVESKIEDFPQISLQDGR-----HYCGGSIYS 62

OY 126 PNKILTNCHVYNTATRSYSAK-GSYVPGMNDSTAVNSANMTFFVPSYINTGASQYD 184
  || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 63 KNVIITAAHCLNRVVAEELRVGVSSYWEHGGSL-----RNISKFOIHESYVEP-TKEYD 116

OY 185 FAVIKTDNTIGNTVGYSIROVTN-----LTG---TTIKISGYPGDKMRSTG-KV 230
  || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 117 VALLKDSLSFNSFTIKAI-ELTNEIPPEYADAIVSGWGETLVPPGIP-DQLRSVDVKI 174

OY 231 SOWEMSGPVTR-----DTNLAY-----YTI--DTFSGNSGSAMLDONOOIVGVHN 274
  || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 175 -----IHREKASRNFYCGSNIKAKMCAVAKGDCQGDGGPLV-VNNLLGVVVS 225

OY 275 AG 276
DB 226 WG 227

RESULT 15
YES9_MYCPN STANDARD; PRT: 569 AA.
ID YES9_MYCPN AC P75324;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothenical lipoprotein MPN459 precursor (H08_orf591).
GN MPN459 OR MP382.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
  pneumoniae.";
Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
  (Potential).
CC -1- SIMILARITY: STRONG, TO THE N-TERMINAL SECTION OF M.PNEUMONIAE
  MPN456 AND M.GENITALIUM MG321.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE000036; AAB96030.1; ALT-INIT.
DR PROSITE; PS00013; PROKAR-LIPOPROTEIN; 1.
KW Hypothenical protein; Lipoprotein; Membrane; Signal;
  Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 569 HYPOTHETICAL LIPOPROTEIN MPN459.
FT LIPID 25 25 N-ACYL DIGLYCERIDE (POTENTIAL).
FT SEQUENCE 569 AA; 63118 MW; 1CE6F077E15B13F3 CRC64;

```

Search completed: June 21, 2002, 18:46:01  
Job time: 220 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2002, 18:38:51 ; Search time 12.97 Seconds  
(without alignments)  
568.738 Million cell updates/sec

Title: US-09-652-743A-12  
Perfect score: 1559  
Sequence: 1 MKVKKLIPSLVFGALVSP.....NGGPKATAAFVEFINYAKAQ 302

ring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
arched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	405.5	26.0	222	1 US-08-090-048-1	Sequence 1, Appl
2	405.5	26.0	222	2 US-08-292-550-1	Sequence 1, Appl
3	405.5	26.0	222	2 US-07-927-661A-1	Sequence 1, Appl
4	184	11.8	344	1 US-08-657-192-3	Sequence 3, Appl
5	184	11.8	344	3 US-08-523-373-5	Sequence 5, Appl
6	184	11.8	392	3 US-08-523-373-6	Sequence 6, Appl
7	184	11.8	532	1 US-08-657-192-9	Sequence 9, Appl
8	184	11.8	532	3 US-08-523-373-7	Sequence 7, Appl
9	184	11.8	537	1 US-08-657-192-15	Sequence 15, Appl
10	183	11.7	213	3 US-08-523-373-22	Sequence 22, Appl
11	183	11.7	214	3 US-08-523-373-23	Sequence 23, Appl
12	183	11.7	215	3 US-08-523-373-24	Sequence 24, Appl
13	182	10.4	241	4 US-08-896-933-32	Sequence 32, Appl
14	182	10.4	241	4 US-09-314-235-32	Sequence 32, Appl
15	134	8.6	246	4 US-08-896-933-33	Sequence 33, Appl
16	134	8.6	246	4 US-09-314-235-33	Sequence 33, Appl
17	129.5	8.3	383	4 US-09-072-384-18	Sequence 18, Appl
18	129.5	8.3	392	4 US-09-072-384-15	Sequence 15, Appl
19	121.5	7.8	256	4 US-09-032-215-27	Sequence 27, Appl
20	121.5	7.8	392	4 US-09-072-384-2	Sequence 2, Appl
21	120	7.7	255	3 US-08-906-769-83	Sequence 83, Appl
22	120	7.7	255	3 US-08-906-616-83	Sequence 83, Appl
23	120	7.7	255	4 US-08-817-795-83	Sequence 83, Appl
24	120	7.7	255	4 US-08-639-075A-83	Sequence 83, Appl
25	120	7.7	255	4 US-09-012-431-83	Sequence 83, Appl
26	120	7.7	255	4 US-09-012-692-83	Sequence 83, Appl
27	120	7.7	255	4 US-08-906-613-83	Sequence 83, Appl

Query Match 26.0%; Score 405.5; DB 1; Length 222;  
Best Local Similarity 39.3%; Pred. No. 3e-29;

28	120	7.7	255	5 PCT-US95-14442A-83	Sequence 83, Appl
29	118	7.6	224	1 US-08-553-516-2	Sequence 2, Appl
30	118	7.6	248	1 US-08-238-130-2	Sequence 2, Appl
31	118	7.6	248	2 US-08-921-426-4	Sequence 4, Appl
32	118	7.6	248	3 US-08-816-915-4	Sequence 4, Appl
33	118	7.6	248	5 PCT-US95-07743-4	Sequence 4, Appl
34	106	6.8	245	3 US-08-906-769-121	Sequence 121, App
35	106	6.8	245	3 US-08-906-616-121	Sequence 121, App
36	106	6.8	245	4 US-08-639-075A-121	Sequence 121, App
37	106	6.8	245	4 US-09-012-431-121	Sequence 121, App
38	106	6.8	245	4 US-09-012-692-121	Sequence 121, App
39	106	6.8	245	4 US-08-906-613-121	Sequence 121, App
40	105	6.7	653	4 US-08-849-602C-27	Sequence 27, Appl
41	105	6.7	653	6 5248599-2	Patent No. 5248599
42	104.5	6.7	255	1 US-08-650-129-5	Sequence 5, Appl
43	104.5	6.7	255	3 US-08-984-417-5	Sequence 5, Appl
44	102	6.5	1536	1 US-08-038-682-2	Sequence 2, Appl
45	102	6.5	1536	1 US-08-302-832-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-090-048-1  
; Sequence 1, Application US/08090048  
; Patent No. 5523237  
; GENERAL INFORMATION:  
; APPLICANT: Budtz, Peter  
; TITLE OF INVENTION: PROTEIN PREPARATIONS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5523237o No. 5523237disk of No. 5523237th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/090,048  
; FILING DATE: 16-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 199/91  
; FILING DATE: 06-FEB-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK92/00036  
; FILING DATE: 06-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Agis, Cheryl H.  
; REGISTRATION NUMBER: 34,086  
; REFERENCE/DOCKET NUMBER: 3396.214-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-867-0298  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 222 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-090-048-1

```

Matches 83; Conservative 36; Mismatches 77; Indels 15; Gaps 5
QY 89 VIGDGRKTVNTRVAPYNSIAYITFGSSCTGTLIAPNKLITNGHCYVNTATRSYSAKG 148
    ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 VIGSDORTVNTAYPYRAIVHISISSIGSCTGWMIGPKTVATAGHCIIYDTSFGSAGTA 61
QY 149 SVTPGMNDSTAVNGSANMTEFYVPSGYINTGASOYDFAVIKTDNIGHTVGYRSIRQVT- 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 TVSPGRNGTSVPYGSVKSTRYFIPSGW-RSGNTNDYGAIEIELSEPIGNVTGVGYFSYTS 120
QY 208 NLTGCTTIKISYPCDKMRSTCKVQWEMSGPVTREDTNLAYTITDTSNGSGSAMLDQNO 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 SLVGTGTTVTSIGYPGDKTAGT----QWHRSGGAISETYKLOYAMDYTGQGGSPVFEQSS 176
QY 268 Q-----TVGVHNAG-YSGNTINGGPKAT 289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 SRTNCSGPCSLAVHTNGVYGGSSYNRGTRIT 207
-----LT 2
    '-292-550-1
    uence 1, Application US/08292550
    cent No. 5863573
; GENERAL INFORMATION:
; APPLICANT: Dambmann, Claus
; APPLICANT: Budtz, Peter
; APPLICANT: Mortensen, Steen B.
; TITLE OF INVENTION: A PROCESS FOR PRODUCING CHEESE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58635730 No. 5863573disk of No. 5863573th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #F.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/292,550
; FILING DATE: 18 AUG 1994
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/927,624
; FILING DATE: 31 AUG 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/0068
; FILING DATE: 8 MAR 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-292-550-1
Query Match 26.0%; Score 405.5; DB 2; Length 222;
Best Local Similarity 39.3%; Pred No. 3e-29;
Matches 83; Conservative 36; Mismatches 77; Indels 15; Gaps 5
QY 89 VIGDGRKTVNTRVAPYNSIAYITFGSSCTGTLIAPNKLITNGHCYVNTATRSYSAKG 148
    ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      2  VIGSDRTRVNTATPYRAIVHSSIGSCTGWMIGPKTVATAGHCIIYDTSSGSPAGTA 61
Qy      149  SVTPGMNDSTAVNGSANMTFFYVPSGYINTGASQYDFAVIKTDTNIGNTVGRSIRQVT- 207
      :| | | | :| | | | | :| | | | :| | | | :| | | | |
Db      62  TVSPGRNGTSYPYGSVKSTRYFIPSCW-RSGNTNIDYGAILESEPIGNTVGYFGSYTTS 120
Qy      208  NLRTGTTIKISYPCDKMRSTGKVSQWEMSGPVTREDTNLAYTYTIDTFSGNSGSAMLQDQ 267
      :| | | | | | | | | | | | | | :| | | | :| | | | | | | |
Db      121  SLVGTTVTSYGPCDKTAGT---QWQHSGP IAISETYKIQYAMDYTGSGSPVFEQS 176
Qy      268  Q-----IVGVHNAG-YSNGTINGGPKAT 289
      :| | | | | | | | | | | | | |
Db      177  SRTNCSGPCSLAVHTNMGVGGSSNRGTRIT 207

RESULT      3
US-07-927-661A-1
; Sequence 1, Application US/07927661A
; Patent No. 5866357
; GENERAL INFORMATION:
; APPLICANT: Dambmann, Claus
; APPLICANT: Mortensen, Steen
; APPLICANT: Budtz, Peter
; APPLICANT: Eriksen, Svend
; TITLE OF INVENTION: Protein Hydrolysates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5866357o No. 5866357disk of No. 5866357th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/927,661A
; FILING DATE: 31-AUG-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3396.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5866357e
US-07-927-661A-1

```

	Query Match	26.0%	Score	405.5;	DB 2;	Length	222;			
	Best Local Similarity	39.3%;	Pred.	No. 3e-29;						
	Matches	83;	Conservative	36;	Mismatches	77;	Indels	15;	Gaps	5;
<hr/>										
Qy	89	VIGDGRKTVNTRVAPYNSIAYITFGSSCTGTLIAPNKILTNCHGVYNATRSYSAGK	148							
		:   :	:   :	:   :	:   :	:   :	:   :			
Dd	2	VIGSDRTRVTNTAYPYRAIVHVISSSIGSCTGMWIGPKTVATAGHCIIYDTSSGSFAGTA	61							
<hr/>										
Qy	149	SVPICMNDSTAVNGSANMTTEFYVPSGYINTCASQYDFAVIKTDINIGTNYCYRSIQVT-	207							
		:     :   :	:   :	:   :	:   :	:   :	:   :			
Dd	62	TVSPGRNGTSYPYGSKSTRIFIPSGW-RSGNTINDYGAIELSEPIGNVTGFVGYSYTT	120							
<hr/>										
Ov	208	NLTGTTIKISGYPGDKMRTCKVKSWEMSGPVRTEDTNLAYYYTIDTFSGNSGSAMLDOHQ	267							

Db 121 SLVGTIVTISGPGDKTACT-----QWQHSQPIAISTETKLYQYAMDTYGGQSGSPVFEQSS 176  
QY 268 Q-----IVGVHNAG-YSNCTINGGPKAT 289  
Db 177 SRTNCSGPCSLAVHTNGVYGGSSYNRGTRIT 207

RESULT 4  
US-08-657-192-3  
; Sequence 3, Application US/08657192  
; Patent No. 5747321  
; GENERAL INFORMATION:  
; APPLICANT: Yabuta, Masayuki  
; APPLICANT: Ohsuye, Kazuhiro  
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8  
; TITLE OF INVENTION: PROTEASES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/657.192  
; FILING DATE: 03-JUN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-170086  
; FILING DATE: 02-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 001560-264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-657-192-3

Query Match 11.8%; Score 184; DB 1; Length 344;  
Best Local Similarity 25.3%; Pred. No. 6.5e-09;  
Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;  
QY 87 KVIIGDGRKVTNTRVAPYNSIAYITF-----GGSSCTGTLIAPNKILTNHCHVYNTATR 142  
Db 124 EVILPNDRHQITDTTNGHYAPVTYIQVEAPTGTGTIASGVVGVGKDTLLTNKHVVDATGHD 183  
QY 143 SYSAGSVYP-GMNDSTAVNG---SANMTEFYVPSGYINTGASQYDFAVIKTDTNIGN-- 196  
Db 184 PHALK--AFPSAINQDNPNGGFTAENITKY-----SGEGDLAIVKFSPEQNKH 231  
QY 197 -----TVGYSRSTQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRETNLAY 248  
Db 232 IGEVVKPATMSNNAETQVNOQ-----NITVTGYPGDKPVAT-----MMESKGIITLKGEAMQ 283  
QY 249 YTIDTFSGNSGAMLDQNOQIIVGVHNAGYSNGTINGGPKATAAFV 293  
Db 284 YDLSTGTGNSGSPVFNKNEKNEVIGIHWGVN-EPNG-----AVFI 322

RESULT 6  
US-08-523-373-6

RESULT 5  
US-08-523-373-5  
; Sequence 5, Application US/08523373  
; Patent No. 6037145  
; GENERAL INFORMATION:  
; APPLICANT: Yabuta, Masayuki  
; APPLICANT: Ohsuye, Kazuhiro  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/523.373  
; FILING DATE: 05-SEP-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-238595  
; FILING DATE: 07-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-296028  
; FILING DATE: 07-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 001560-251  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-523-373-5

Query Match 11.8%; Score 184; DB 3; Length 344;  
Best Local Similarity 25.3%; Pred. No. 6.5e-09;  
Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;  
QY 87 KVIIGDGRKVTNTRVAPYNSIAYITF-----GGSSCTGTLIAPNKILTNHCHVYNTATR 142  
Db 124 EVILPNDRHQITDTTNGHYAPVTYIQVEAPTGTGTIASGVVGVGKDTLLTNKHVVDATGHD 183  
QY 143 SYSAGSVYP-GMNDSTAVNG---SANMTEFYVPSGYINTGASQYDFAVIKTDTNIGN-- 196  
Db 184 PHALK--AFPSAINQDNPNGGFTAENITKY-----SGEGDLAIVKFSPEQNKH 231  
QY 197 -----TVGYSRSTQVNTLTGTTIKISGYPGDKMRSTGKVSQWEMSGPVTRETNLAY 248  
Db 232 IGEVVKPATMSNNAETQVNOQ-----NITVTGYPGDKPVAT-----MMESKGIITLKGEAMQ 283  
QY 249 YTIDTFSGNSGAMLDQNOQIIVGVHNAGYSNGTINGGPKATAAFV 293  
Db 284 YDLSTGTGNSGSPVFNKNEKNEVIGIHWGVN-EPNG-----AVFI 322

; Sequence 6, Application US/08523373  
; Patent No. 6037145

## GENERAL INFORMATION:

; APPLICANT: Yabuta, Masayuki  
; APPLICANT: Ohsuye, Kazuhiro  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/523,373  
; FILING DATE: 05-SEP-1995  
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-238595  
; FILING DATE: 07-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-296028  
; FILING DATE: 07-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 001560-251  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021

; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 392 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-523-373-6

Query Match 11.8%; Score 184; DB 3; Length 392;

Best Local Similarity 25.3%; Pred. No. 7.7e-09;

Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;

87 KVVIGDGRKTVNTRVAPYNSIAYITF-----GGSSCTGTLIAPNKLTHNGHCYVNTATR 142

124 EVILPNDRHQITDTHNGHYAPVTYIQVEAPTGTFTIASGVVVGKDTLLTNKHVVVDATHGD 183

143 SYSAGSVYP-GMNDSTAVNG---SANTMEFYVPYSGYINTGASQYDFAVIKTDNIGN-- 196

184 PHALK--AFPSAINQDNPNGGTAENITKY-----SGEGDLAIIVKFSPEONKH 231

197 -----TVGYSIRQVTLNLTGTTIKISGPGDKMRSTCKVQSWEMSGPVTREDTNLAY 248

232 IGEVVKPATMSNNAETQVNO-----NITVTGYPGDKPVAT-----MWESKGIITLKGEAMQ 283

249 YTIDTFSNGSAMLQDNOQOIVGVHNGYNSGTINGGPKATAAFV 293

284 YDLSTTGGSGSPVNEKNEVIGIHWGGVFN-EFNG-----AVFI 322

RESULT 7

US-08-657-192-9

; Sequence 9, Application US/08657192

; Patent No. 5747321

## GENERAL INFORMATION:

; APPLICANT: YABUTA, Masayuki

; Sequence 6, Application US/08523373

## GENERAL INFORMATION:

; APPLICANT: Yabuta, Masayuki  
; APPLICANT: Ohsuye, Kazuhiro  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/523,373  
; FILING DATE: 05-SEP-1995  
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-238595  
; FILING DATE: 07-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-296028  
; FILING DATE: 07-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 001560-264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 532 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-657-192-9

Query Match 11.8%; Score 184; DB 1; Length 532;

Best Local Similarity 25.3%; Pred. No. 1.2e-08;

Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;

87 KVVIGDGRKTVNTRVAPYNSIAYITF-----GGSSCTGTLIAPNKLTHNGHCYVNTATR 142

124 EVILPNDRHQITDTHNGHYAPVTYIQVEAPTGTFTIASGVVVGKDTLLTNKHVVVDATHGD 183

143 SYSAGSVYP-GMNDSTAVNG---SANTMEFYVPYSGYINTGASQYDFAVIKTDNIGN-- 196

184 PHALK--AFPSAINQDNPNGGTAENITKY-----SGEGDLAIIVKFSPEONKH 231

197 -----TVGYSIRQVTLNLTGTTIKISGPGDKMRSTCKVQSWEMSGPVTREDTNLAY 248

232 IGEVVKPATMSNNAETQVNO-----NITVTGYPGDKPVAT-----MWESKGIITLKGEAMQ 283

249 YTIDTFSNGSAMLQDNOQOIVGVHNGYNSGTINGGPKATAAFV 293

284 YDLSTTGGSGSPVNEKNEVIGIHWGGVFN-EFNG-----AVFI 322

RESULT 8

US-08-523-373-7

; Sequence 7, Application US/08523373

; Patent No. 6037145

## GENERAL INFORMATION:

; APPLICANT: Yabuta, Masayuki

; APPLICANT: Ohsuye, Kazuhiro

; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

; STREET: 699 Prince Street

CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-7

Query Match 11.8%; Score 184; DB 3; Length 532;  
Best Local Similarity 25.3%; Pred. No. 1.2e-08;  
Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;  
QY 87 KVVIGDGRKVTNTRVAPYNSIAYITF----GGSCCTGLIAPNPKILTNHCHVYNTAPR 142  
DB 124 EVILPNDRHQITDTTNHGYAPVTYIQVEAPGTFTIASGVVYVKDITLLTNKHVVDTAGHD 183  
QY 143 SYSAGSVYP-GMNDSTAVNG---SANMTEFYVPSGYINTGASQYDFAVIKTDTNIGN-- 196  
DB 184 PHALK--APPSAINQDNYPNGGFTAENITKY-----SGEGDLAIVKFSPEQNKH 231  
QY 197 -----TVGRSIRQVNTLTGTTIKISGYFGDKMRSTGKVSOWEMSGPVTREDTNLAY 248  
DB 232 ICEVVKPATMSNNAETOVNO----NITVTGYFGDKRPVAT-----AVFI 322  
QY 249 YTIIDTFSGNSGAMLDONQOIVGVHNAGYSNGTINGGPKATAAFV 293  
DB 284 YDLSTTGGNSGSPVFNKNEVIGIHWGGVYPN-EFNG-----AVFI 322

RESULT 9  
US-08-657-192-15  
Sequence 15, Application US/08657192  
Patent No. 5747321  
GENERAL INFORMATION:  
APPLICANT: YABUTA, Masayuki  
APPLICANT: OHSUYE, Kazuhiro  
TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8  
TITLE OF INVENTION: PROTEASES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States

ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/657,192  
FILING DATE: 03-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-170086  
FILING DATE: 02-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-264  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 537 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-657-192-15

Query Match 11.8%; Score 184; DB 1; Length 537;  
Best Local Similarity 25.3%; Pred. No. 1.2e-08;  
Matches 57; Conservative 41; Mismatches 83; Indels 44; Gaps 10;  
QY 87 KVVIGDGRKVTNTRVAPYNSIAYITF----GGSCCTGLIAPNPKILTNHCHVYNTATR 142  
DB 124 EVILPNDRHQITDTTNHGYAPVTYIQVEAPGTFTIASGVVYVKDITLLTNKHVVDTAGHD 183  
QY 143 SYSAGSVYP-GMNDSTAVNG---SANMTEFYVPSGYINTGASQYDFAVIKTDTNIGN-- 196  
DB 184 PHALK--APPSAINQDNYPNGGFTAENITKY-----SGEGDLAIVKFSPEQNKH 231  
QY 197 -----TVGRSIRQVNTLTGTTIKISGYFGDKMRSTGKVSOWEMSGPVTREDTNLAY 248  
DB 232 ICEVVKPATMSNNAETOVNO----NITVTGYFGDKRPVAT-----AVFI 322  
QY 249 YTIIDTFSGNSGAMLDONQOIVGVHNAGYSNGTINGGPKATAAFV 293  
DB 284 YDLSTTGGNSGSPVFNKNEVIGIHWGGVYPN-EFNG-----AVFI 322

RESULT 10  
US-08-523-373-22  
Sequence 22, Application US/08523373  
Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohsuye, Kazuhiro  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:



```

RESULT 11
US-08-523-373-23
: Sequence 23, Application US/08523373
: Patent No. 6037145
:
: ENERAL INFORMATION:
: APPLICANT: Yabuta, Masayuki
: APPLICANT: Ohsuye, Kazuhiro
: TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
:
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
: STREET: 699 Prince Street
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22314-3187
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/523,373
: FILING DATE: 05-SEP-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:

```

```

RESULT 12
US-08-523-373-24
; Sequence 24, Application US/08523373
; Patent No. 6037145
;
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohsuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028

```

FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Neuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELEPHONE: 703-836-6620  
TELEPHONE: 703-836-2021  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 215 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
78-523-373-24

Query Match 11.78; Score 183; DB 3; Length 215;  
Best Local Similarity 23.48; Pred. No. 4.le-09;  
Matches 57; Conservative 40; Mismatches 83; Indels 44; Gaps 10;  
Qy 88 VVIGDGRKTKVNTVAQVNSIATITF---GGSSCTGTLIAPNKLITNGHCYVNTATRS 143  
Db 1 VILPNDRHQITDTHNGHAPYTIQVEAPTGTFTIASGVVYVCKDTLLTNKHHVDATHGDP 60  
Qy 144 YSAKGSVYP-GMNDSTAVNG---SANTFEYVPSGYINTGASQYDFAVIKTDTNIGN--- 196  
Db 61 HALK--AFPSAINODNYPNGGTAENITKY-----SGEGDLAIVKFSPEQNKHI 108  
Qy 197 -----TVGRSIRQVNTLGTITKISGYPGDKMRSTGKVSOWEMSPGVTREDTNLAY 249  
Db 109 GEVVKPATMSNAETQVQV---NITVTGFCDKPVAT---MMESKGIITLKGAEAMQY 160  
Qy 250 TIDTFSGNSGAMLQOQOIVGVHNGYSNGTINGGPKATAAFV 293  
Db 161 DLSTTGGNSGSPVFNKNEVIGIHWGGVNP-EFNG-----AVFI 198

RESULT 13  
US-08-896-933-32  
; Sequence 32, Application US/08896933  
; Patent No. 6221351  
; GENERAL INFORMATION:  
; APPLICANT: Terman, David S.  
; TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,  
; TITLE OF INVENTION: SUPERANTIGENS, AND RELATED COMPOUNDS  
; FILE REFERENCE: 09629/005002  
; CURRENT APPLICATION NUMBER: US/08/896,933  
; CURRENT FILING DATE: 1997-07-18  
; EARLIER APPLICATION NUMBER: 08/252,978  
; EARLIER FILING DATE: 1994-06-02  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 32  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-08-896-933-32

Query Match 10.48; Score 162; DB 4; Length 241;  
Best Local Similarity 30.08; Pred. No. 3.9e-07;  
Matches 62; Conservative 26; Mismatches 93; Indels 26; Gaps 9;  
Qy 105 PYNSTAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAKGSVYPCMDNDSTAVNGSA 164  
Db 42 PYNTIGNVFKGTSATGLICKNTVLTNRHIA--KFANGDPSKVSFRSINTDDNGNTET 99  
Qy 165 NMTEFYVPSGYINTGASQYDFAVI---KTDN---IGNTVGYSIRQVNTNL-TGTTIKISG 218  
Db 100 PYGEYEVKEILQEPFGAGVDLALRLKPDQNGVSLGDKISPAKIGTSDNLKDGKLELIG 159  
Qy 219 YPGDKMRSTGKVSOWEMSG-PVTREDTNLAYTITDFTSCNSGAMLQOQOIVGVHNGY 277  
Db 160 YPFDH-----KVNQMRSEIELTILSRGLRYGF-TVPCNSGSGIFNSGELVGIHSSKV 213  
Qy 278 SN-----GTINGGPKATAAFVEFINYAK 300  
Db 214 SHLDREHQINYG-----VGIGNYVK 233

RESULT 15  
US-08-896-933-33  
; Sequence 33, Application US/08896933  
; Patent No. 6221351  
; GENERAL INFORMATION:  
; APPLICANT: Terman, David S.  
; TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,  
; TITLE OF INVENTION: SUPERANTIGENS, AND RELATED COMPOUNDS  
; FILE REFERENCE: 09629/005002  
; CURRENT APPLICATION NUMBER: US/08/896,933  
; CURRENT FILING DATE: 1997-07-18

Qy 219 YPGDKMRSTGKVSOWEMSG-PVTREDTNLAYTITDFTSCNSGAMLQOQOIVGVHNGY 277  
Db 160 YPFDH-----KVNQMRSEIELTILSRGLRYGF-TVPCNSGSGIFNSGELVGIHSSKV 213  
Qy 278 SN-----GTINGGPKATAAFVEFINYAK 300  
Db 214 SHLDREHQINYG-----VGIGNYVK 233  
RESULT 14  
US-09-314-235-32  
; Sequence 32, Application US/09314235  
; Patent No. 6338845  
; GENERAL INFORMATION:  
; APPLICANT: Terman, David S.  
; TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,  
; TITLE OF INVENTION: SUPERANTIGENS, AND RELATED COMPOUNDS  
; FILE REFERENCE: 09629/005004  
; CURRENT APPLICATION NUMBER: US/09/314,235  
; CURRENT FILING DATE: 1999-05-18  
; EARLIER APPLICATION NUMBER: 08/896,933  
; EARLIER FILING DATE: 1997-07-18  
; EARLIER APPLICATION NUMBER: 08/252,978  
; EARLIER FILING DATE: 1994-06-02  
; EARLIER APPLICATION NUMBER: 07/891,718  
; EARLIER FILING DATE: 1992-06-01  
; EARLIER APPLICATION NUMBER: US91/00342  
; EARLIER FILING DATE: 1991-01-17  
; EARLIER APPLICATION NUMBER: 07/466,577  
; EARLIER FILING DATE: 1990-01-17  
; EARLIER APPLICATION NUMBER: 07/416,530  
; EARLIER FILING DATE: 1989-10-03  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 32  
; LENGTH: 241  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-314-235-32

Query Match 10.48; Score 162; DB 4; Length 241;  
Best Local Similarity 30.08; Pred. No. 3.9e-07;  
Matches 62; Conservative 26; Mismatches 93; Indels 26; Gaps 9;  
Qy 105 PYNSTAYITFGSSCTGLIAPNKLITNGHCYVNTATRSYSAKGSVYPCMDNDSTAVNGSA 164  
Db 42 PYNTIGNVFKGTSATGLICKNTVLTNRHIA--KFANGDPSKVSFRSINTDDNGNTET 99  
Qy 165 NMTEFYVPSGYINTGASQYDFAVI---KTDN---IGNTVGYSIRQVNTNL-TGTTIKISG 218  
Db 100 PYGEYEVKEILQEPFGAGVDLALRLKPDQNGVSLGDKISPAKIGTSDNLKDGKLELIG 159  
Qy 219 YPGDKMRSTGKVSOWEMSG-PVTREDTNLAYTITDFTSCNSGAMLQOQOIVGVHNGY 277  
Db 160 YPFDH-----KVNQMRSEIELTILSRGLRYGF-TVPCNSGSGIFNSGELVGIHSSKV 213  
Qy 278 SN-----GTINGGPKATAAFVEFINYAK 300  
Db 214 SHLDREHQINYG-----VGIGNYVK 233

; EARLIER APPLICATION NUMBER: 08/252,978
; EARLIER FILING DATE: 1994-06-02
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-08-896-933-33

Query Match	8.6%	Score 134;	DB 4;	Length 246;
Best Local Similarity	28.4%	Pred. No. 0.00014;		
Matches	58;	Conservative 27;	Mismatches 83;	Indels 36;
				Gaps 10;

  

QY	96	TKVTNTRVAPYNSIAIYITFGSS-CTGTLIAPNKILTNHCHVYNTATRSYSAKGSVYPGM	154
DB	26	THITDNARSPYNSVGTVFVKGSTLATGVLIGKNTIVTNYH-VAREAAKNPS-----NI	77
		155 NDSTAVNGSANMTEFYVPSGYINT---GASQY-----DFAVIKTDTN-IGNTVGYRSIROV	206
		78 IFTPAQNRAEKNEFPTYGKFEAEIKESPYGGGLDLAIKLPNEKESAG--DLIQP	135
QY	207	TNL-----TGTTIKISGYPGDKMRSTGKVSQWENSGPVTREDTNLAYTTIDTFSGNSG	259
DB	136	ANIPRHDIOKDKYSLGYPYNSAYSLSYQSIEM-----FNDSOYFGYTEVGNNG	187
QY	260	SAMLDQNOQIIVGVHNAGYNGTIN	283
DB	188	SGIFNLKGELIGIHSG--KGGQMN	209

Search completed: June 21, 2002, 18:42:16
Job time: 205 sec